

AD-A211 117

LASER PHYSICS AND LASER TECHNIQUES

Final Technical Report to
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
under Contract No. F49620-86-K-0013

for the period
1 March 1986 — 28 February 1989

Principal Investigator
A. E. Siegman
Professor of Electrical Engineering

Edward L. Ginzton Laboratory
W.W. Hansen Laboratories of Physics
Stanford University
Stanford, California 94305

June 1989

DTIC
ELECTE
AUG 10 1989
S B D

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT <i>Approved for Public Release; Distribution Unlimited</i>		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-89-1073		
4. PERFORMING ORGANIZATION REPORT NUMBER(S)					
6a. NAME OF PERFORMING ORGANIZATION Stanford University		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research		
6c. ADDRESS (City, State and ZIP Code) Edward L. Ginzton Laboratory Stanford, California 94305			7b. ADDRESS (City, State and ZIP Code) Directorate of Physical and Geophysical Sciences, Bolling AFB, DC 20332		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION <i>Same as 7a</i>		8b. OFFICE SYMBOL (If applicable) <i>NP</i>	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Contract F49620-86-K-0013		
8c. ADDRESS (City, State and ZIP Code) <i>Same as 7b</i>		10. SOURCE OF FUNDING NOS.			
		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2301	TASK NO. A1	WORK UNIT NO.
11. TITLE (Include Security Classification) Laser Physics and Laser Techniques (unclassified)					
12. PERSONAL AUTHOR(S) <i>Dr. A. C. Siegman</i>					
13a. TYPE OF REPORT Final Report		13b. TIME COVERED FROM 3/1/86 TO 2/28/89		14. DATE OF REPORT (Yr., Mo., Day) June 1989	
15. PAGE COUNT 165					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.	optics; lasers; nonlinear optics; unstable resonators; excess biuse; Spontaneous emission; ultrafast photo-detectors; optical Kerr effect.		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) We report on accomplishments in several different areas of Laser Physics and Laser Techniques, including ultrafast physical measurements; the development of a new subpicosecond time-response photodetector; identification and analysis of an important new quantum noise limit for unstable laser oscillators; and useful advances in stable and unstable laser resonator theory.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL <i>Dr. Howard Schlossberg</i>		22b. TELEPHONE NUMBER (Include Area Code) 202/767-4906		22c. OFFICE SYMBOL <i>NP</i>	

89 8 10 079

I. INTRODUCTION

This is a final report and summary of accomplishments on Air Force Office of Scientific Research Contract No. F49620-86-K-0013 by the research group of Professor A. E. Siegman in the Edward L. Ginzton Laboratory at Stanford University. It reports accomplishments in several different areas of Laser Physics and Laser Techniques, including ultrafast physical measurements; the development of a new subpicosecond time-response photodetector; identification and analysis of an important new quantum noise limit for unstable laser oscillators; and useful advances in stable and unstable laser resonator theory.

Since the primary reporting mechanism for an academic research program such as this is through scholarly publications, we will summarize the results of this program in a number of sections dealing with different areas of research, giving in each case an annotated listing of the research publications prepared under AFOSR support in each of these areas. Reprints or abstracts of all these publications are then attached as Appendices to this report.

II. LASER-INDUCED SURFACE GRATINGS AND QUANTUM-WELL LIFETIMES

We note first two areas of research in which the research work was primarily done on earlier AFOSR contracts, but with results published during the period of the present contract. One of these areas is the very interesting topic of laser-induced surface ripples, or "stimulated Wood's anomalies," a topic on which we had made earlier experimental and theoretical studies, and on which we were asked to prepare an invited review paper:

A.E. Siegman and P. M. Fauchet, "Stimulated Wood's anomalies on laser-illuminated surfaces" (invited paper), *IEEE J. Quantum Electron.* QE-22, 1384-1403 (August 1986).

These spontaneously growing ripples are of importance in laser damage studies, in laser photodeposition on semiconductors, in grating coupling to photodetectors, and in surface physics studies using lasers. Hence, we hope this review will be useful to workers in these areas. Dr. Philippe Fauchet, who did much of this work, is now continuing many other laser studies of semiconductor phenomena as a faculty



For	<input checked="checked" type="checkbox"/>
I	<input type="checkbox"/>
d	<input type="checkbox"/>
tion	

on/

Availability Codes	
Dist	Avail and/or Special
A-1	

member at Princeton University.

The second topic is the Ph.D. dissertation work of Dr. Julie Fouquet on measuring lifetimes in quantum well structures. This work has since led to two further publications based on her Stanford experimental data, namely

J. E. Fouquet, "Recombination dynamics in microstructures," *SPIE Proc. Vol. 793: Ultrafast Laser Probe Phenomena in Bulk and Microstructure Semiconductors*, Bay Point Florida, 25-26 May 1987, pp. 30—66.

J.E. Fouquet and R. D. Burnham, "Recombination dynamics in GaAs/Al_xGa_{1-x}As quantum well structures," *IEEE J. Quantum Electron.* QE-22, 1799-1810 (September 1986).

Dr. Fouquet is now continuing her research on the electrooptical properties of III-V compounds and quantum-well structures at H-P Laboratories.

III. FEMTOSECOND NONLINEAR OPTICAL MEASUREMENTS USING TUNABLE LASER-INDUCED GRATINGS

One of our major areas of effort over the past six or more years has been the development and application of a novel method for making ultrafast (picosecond and femtosecond time scale) optical measurements using tunable laser-induced gratings. In essence, we make these ultrafast measurements of the nonlinear optical responses of different kinds of materials in the frequency domain, using cw or long-pulse but tunable laser beams, rather than in the time domain using ultrashort optical pulses. The time-domain and frequency-domain approaches both have their own advantages and disadvantages; we have been pioneers in developing and exploiting the frequency-domain tunable-grating approach.

The basic tunable laser-induced grating technique was first developed and put to work through the Ph.D. work of Dr. Rick Trebino under earlier AFOSR support. Much of this work is reported in an invited review paper

Rick Trebino, Charles E. Barker, and A. E. Siegman, "Tunable-laser-induced gratings for the measurement of ultrafast phenomena," (invited paper), *IEEE J. Quantum Electron.* QE-22, 1413-1430 (August 1986).

and two other journal publications

Rick Trebino, E. K. Gustafson, and A. E. Siegman, "Fourth-order partial-coherence effects in the formation of integrated-intensity gratings with pulsed light sources," *J. Opt. Soc. Am. B* 3, 1295-1304 (October 1986).

Rick Trebino and A. E. Siegman, "Frequency bandwidths in nondegenerate N-wave-mixing interactions and induced-grating geometries," *Optics Commun.* 56, 297-302 (December 1986).

which were published under the present contract. This research also led to several developments in tunable dye laser technology, one of which was reported during the present contract period:

Rick Trebino, Charles E. Barker, and A. E. Siegman, "Achromatic N-prism beam expanders: Optimal configurations II," *SPIE Proceedings: Southwest Conference on Optics* (1986).

Dr. Trebino is now applying similar laser-induced grating methods to combustion and other diagnostics problems at Sandia Laboratories in Livermore, California.

The tunable grating project was continued during the present contract as the Ph.D. work of Dr. Charles Barker. Barker obtained in particular important new results on the optical Kerr effect or nonlinear refractive index in CS₂, as reported in the publications:

Rick Trebino, Charles Barker, and Adnah G. Kostenbauder, "Rise-fall ambiguities and their removal from frequency-domain ultrafast-measurement techniques," *Ultrafast Phenomena VI*, in press (Springer-Verlag, 1988); and also "Rise-fall ambiguities and their removal from frequency-domain nonlinear-optical techniques," *Optics. Lett.* 13, 865-867 (October 1988).

Charles E. Barker, Rick Trebino, A. G. Kostenbauder, and A. E. Siegman, "Frequency-domain observation of the ultrafast inertial response of the optical Kerr effect in CS₂," *J. Chem. Phys.*, submitted for publication (March 1989).

Besides obtaining data on different tensor components of the third-order nonlinear susceptibility tensor in CS₂ (a special capability of the tunable grating method), Barker has now extended the transient grating method well into the femtosecond regime, and confirmed the inertial nature of the optical Kerr response $\Delta n(t)$ in CS₂.

Figure 1 summarizes some typical results from this work. On the left in each row are typical data from our tunable grating method, showing diffracted probe light

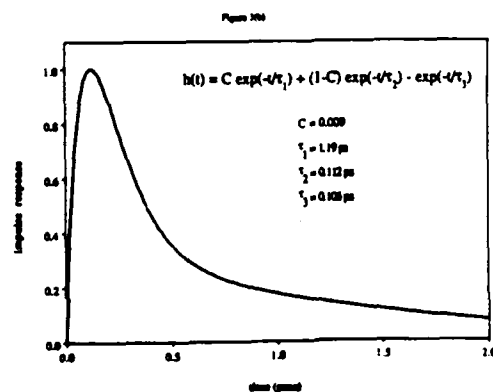
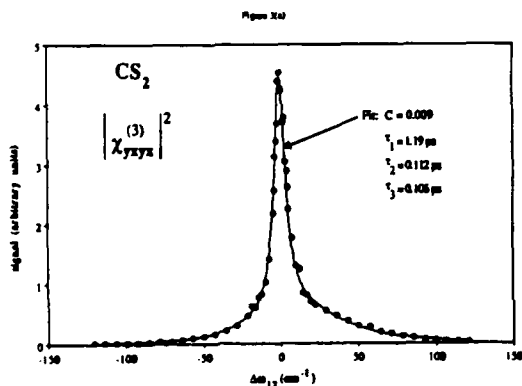
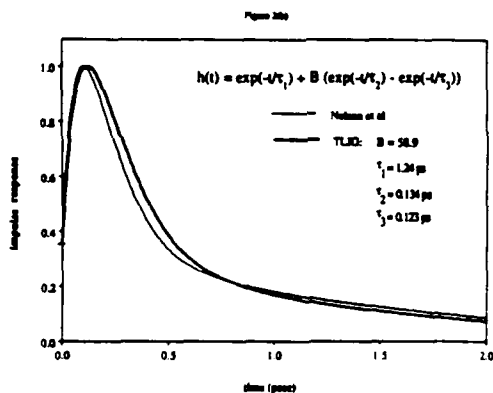
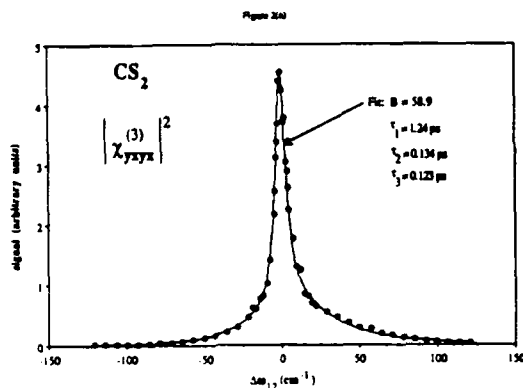
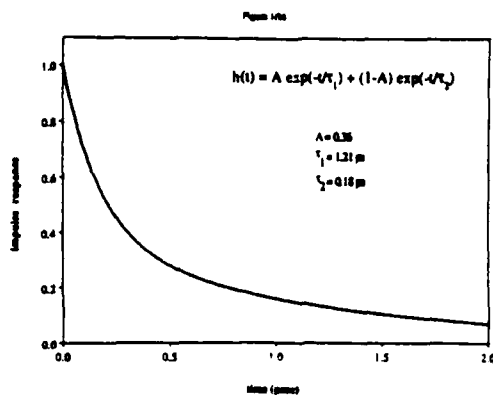
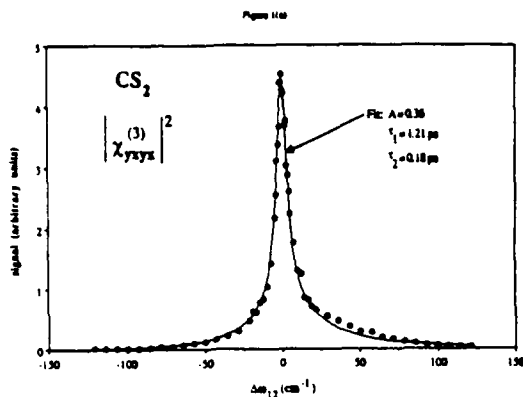


Figure 1. A typical set of data from our tunable laser-induced grating experiments on the $yxxy$ tensor component of the optical Kerr effect in CS₂. The same data points are fit with three successively more complex expressions for the impulse response of the nonlinear index change $\Delta n(t)$ as shown on the righthand side. The two inertial response curves (with a finite rise time) give a much more accurate fit to the data.

intensity versus the difference frequency between two excitation lasers that form the moving gratings. In each of the three rows the same data are fit with successively more complicated models for the impulse response $\Delta n(t)$ of the optical Kerr effect, with the impulse response $\Delta n(t)$ shown on the righthand side. Note that our method measures in essence the Fourier transform squared of the impulse response $\Delta n(t)$ of the CS_2 sample. The exact form for $\Delta n(t)$ is not known theoretically and, in common with all other workers, we must attempt to find reasonable analytic expressions which best fit the measured data. We have developed and used for our data analysis a simplex fitting method considerably more sophisticated than that used by other researchers to date.

The inertial form of the optical Kerr response—that is, the fact that $\Delta n(t)$ has a finite turn-on time, with a response time of ~ 100 fsec, as well as finite decay times on the order of 100 fsec and 1-2 psec—was first observed only recently using fsec pulse techniques by Nelson at MIT and Kenney-Wallace at Toronto. Our frequency-domain results, as illustrated in Figure 1, fully confirm these time-domain results. Indeed our fits make clear that a finite rise time or turn-on time is essential to obtain a good fit to the observed data. In physical terms, it requires ~ 100 fsec for the CS_2 molecules to respond to an applied optical E field with the molecular deformation and rotation necessary to produce the $\Delta n(t)$.

It is also clear from our measurements and our extensive curve-fitting efforts, that while a finite inertial rise time is essential, none of the proposed analytical models is sufficient to give a fully accurate description of $\Delta n(t)$. Slightly different inertial models give slightly better or worse statistical fits, all of them fairly good, as shown in the lower two plots of Figure 1; but none of them fully meet the statistical criteria for a completely satisfactory fit. The various multiexponential analytical models suggested by the femtosecond pulse experimenters, therefore, must be accepted only as approximations, and not as definitive descriptions of the time constants in the CS_2 system.

Finally, besides the tunable grating experiments, we carried out and reported a collaborative experiment on the electronic and nuclear components third-order susceptibility in an interesting polysilane material, as reported in

Daniel J. McGraw, A. E. Siegman, G. M. Wallraff, and R. D. Miller, "Resolution of the nuclear and electronic contributions to the optical nonlinearity in polysilanes," *Appl. Phys. Lett.* **54**, 1713-1715 (1 May 1989).

This work was done during a postdoctoral stay by Dr. Dan McGraw, who has now become an Assistant Professor of Physics at the University of New Mexico

IV. HIGH-SPEED DIFFUSION-DRIVEN PHOTODETECTORS

Several years ago, as a graduate student on this AFOSR program, Ad Kostenbauder proposed a very novel concept for a high-speed diffusion-driven photovoltaic (unbiased) photodetector, in which the photo response was produced by the differential motion of holes and electrons in an asymmetric carrier grating produced in any suitable semiconductor. The interesting and very useful properties of this detector have since been fully verified in experiments reported in two publications, namely

Adnah G. Kostenbauder, "High speed diffusion-driven photodetector," *Appl. Phys. Lett.* **51**, 1129-1131 (12 October 1987).

Adnah Kostenbauder, S. J. B. Yoo, and A. E. Siegman, "A fast diffusion-driven photodetector: Theory and experiment," *IEEE J. Quantum Electron.* **24**, 240-244 (February 1988).

The first of these papers reported measured response times down to about 1 nsec in a very simple "edge-type" silicon structure. The second publication extended the measured response time down to ~40 psec in an almost equally simple "gap-type" silicon detector.

The most recent form of this detector which we have now developed, with measured response times under 2 psec, employs the still quite simple structure shown in Figure 2. In essence the detector consists only of a coplanar Al stripline, with two ~5 μm wide stripes spaced ~10 μm apart, on the surface of a 500 Å thick layer of Si fabricated by conventional silicon-on-sapphire techniques on a sapphire substrate. The detector is illuminated by an ultrafast light pulse which is incident on one of the Al stripes from underneath, so as to form a single half-period of a vertical standing-wave grating in the Si layer under the Al stripe. The femtosecond

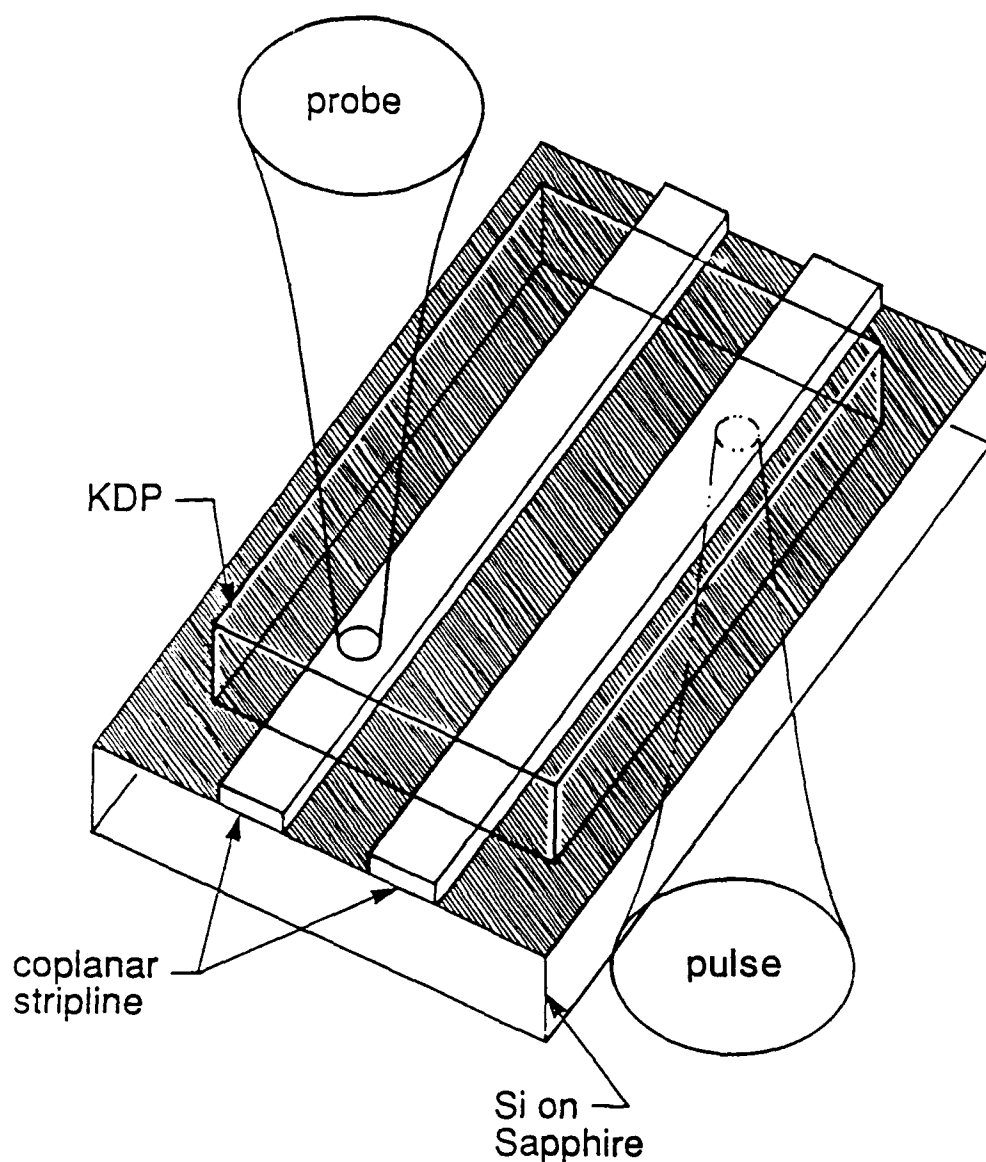


Figure 2. Structure of the subpicosecond diffusion-driven photodetector. The coplanar stripline is deposited on a 500Å thick silicon-on-sapphire layer, and capped by a KDP crystal. The optical signal pulse comes from below, is reflected from one of the aluminum striplines, and excites an electrical impulse traveling down the coplanar line from upper right to lower left. The probe beam samples the traveling electrical pulse fields electrooptically in the KDP crystal a short distance down the line.

decay of this grating excites through the Demer effect an ultrashort electrical pulse on the coplanar transmission line; and this pulse is probed electrooptically by a delayed optical probe pulse coming from above through the KDP slab at a point a short distance down the transmission line.

Figure 3 shows on the same time scale the electrooptically sampled response of this detector using ~ 2 psec long excite and probe pulses, and also the purely optical autocorrelation between the same pulses using optical second harmonic generation. The essential result is that on this time scale the diffusion-driven photodevice shows no observable broadening in time beyond that due to the incident mode-locked laser pulsewidth of ~ 2 psec. The conclusion is that the photodetector itself must have a response time certainly not longer than 700 fsec, and probably closer to ~ 300 fsec. This work has only recently been completed, and will be submitted for publication shortly.

In addition, we have recently done another experiment in which the bipolar response of one of the gap-type photodetectors was combined with the properties of the two-mode optical fibers developed by Professor B. Y. Kim in this laboratory to produce a nsec response time bipolar fiber modulation and detection system, as reported in

Paul Wysocki, A. G. Kostenabuder, B. Y. Kim, and A. E. Siegman, "Bipolar optical modulation and demodulation using a dual-mode fiber and a fast diffusion-driven photodetector," *IEEE J. Lightwave Technol.*, accepted for publication (April 1989).

This publication is now in press.

V. EXCESS QUANTUM NOISE FLUCTUATIONS IN UNSTABLE-RESONATOR LASERS

K. Petermann in Germany first predicted in 1979 the possible existence of a so-called "excess spontaneous emission" or "excess quantum noise effect," corresponding to a spontaneous emission level greater than one extra photon per cavity mode, which he predicted would occur in gain-guided semiconductor lasers. This prediction was initially controversial since the same analysis appeared to predict violations of fundamental thermodynamic considerations in, for example,

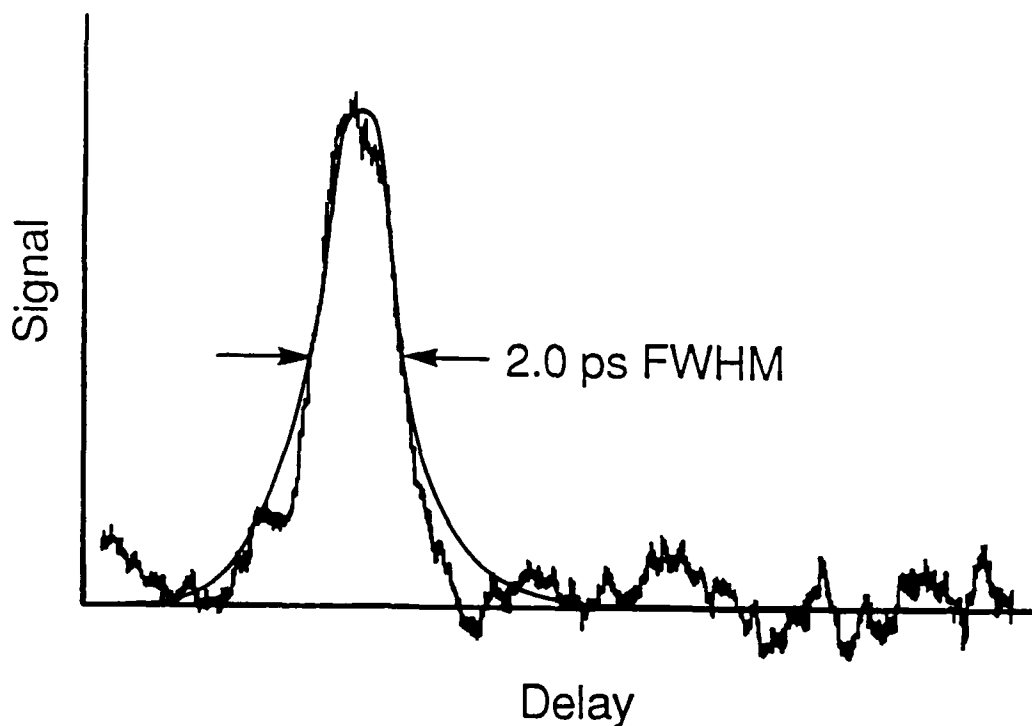


Figure 3. Measured performance of the silicon-on-sapphire diffusion-driven photodetector. The photoexcited and electro-optically sampled autocorrelation response of the photodetector (jagged trace) is not observably broadened as compared to the direct optical second harmonic autocorrelation of the 2 psec optical pulse (smooth trace), indicating that the time response of the photodetector must be substantially faster than 1 picosecond.

equivalent passive loss-guided systems. These difficulties were cleared up in 1985 by Haus and Kawakami, who noted that correlation between noise emission into different transverse modes would resolve at least the fundamental thermodynamic difficulties in Petermann's prediction.

From our experience with unstable resonators and the orthogonality properties of optical resonator modes, we soon realized that essentially the same excess spontaneous emission rate or excess quantum noise emission should in fact be present in all types of laser cavities, especially in any kind of geometrically unstable resonator, rather than just in the gain-guided systems considered by Petermann and a number of subsequent authors. One can show that these excess noise properties arise in fact entirely from the non-power-orthogonal or biorthogonal nature of the modes in these optical cavities. They have nothing directly to do with the presence or absence of any gain-guiding effects.

We therefore carried out during this contract period a complete and detailed fundamental derivation of the excess spontaneous emission process as it applies to both laser amplifiers and laser oscillators. This work has now been reported in two Physical Review articles

A. E. Siegman, "Excess spontaneous emission in nonhermitian optical systems. I. Laser amplifiers," *Phys. Rev. A.* **39**, 1253-1263 (1 February 1989).

A. E. Siegman, "Excess spontaneous emission in nonhermitian optical systems. II. Laser oscillators," *Phys. Rev. A.* **39**, 1264-1268 (1 February 1989).

These two papers derive general formulas for the excess noise factors and noise fluctuations in arbitrary types of open-sided laser cavities having uniform gain, with the noise factors expressed in terms of the overlap integrals and biorthogality factors of the relevant cavity modes. Calculating numerical values for the excess noise coefficients in any specific cavity still requires, however, detailed knowledge of the resonator eigenmodes and adjoint modes for that cavity. We have also carried out therefore a detailed calculation of the noise factors for the analytically solvable class of geometrically stable or unstable variable-reflectivity-mirror (VRM) laser cavities; and these results will be published in

A. E. Siegman, Paul L. Mussche, and Jean-luc Doumont, "Excess spontaneous emission factor in variable-reflectivity-mirror lasers," *IEEE J. Quantum Electron.*, accepted for publication (March 1989).

Figure 4 shows as an example the cavity power loss per round trip and the excess noise factor for one of these cavities having a certain gaussian mirror reflectivity as a function of the geometrical magnification M of the cavity optics. An important feature of these results is that for rather moderate degrees of geometrical instability, where the power loss per bounce (or fractional output coupling) is still quite reasonable and where a practical laser might operate, the excess noise factor can still become very large, on the order 100 to 1000 times larger than in stable cavities.

Figure 5 shows similar data for gaussian VRM cavities with different values of geometrical magnification M , as a function of gaussian aperture size or Fresnel number. Note that planar resonator modes ($M = 1$) have an excess noise factor $K_p = 2$ except for very small Fresnel numbers.

This excess noise factor is of fundamental importance in determining the quantum limit for the spectral broadening or the so-called Schawlow-Townes limit in ultrastable laser oscillators, as well as the injection seeding behavior of high-power laser cavities. We are now planning an experimental test of this excess noise factor as one of the major projects under our continuing program.

VI. LASER RESONATOR THEORY

Laser resonator theory is by now a relatively mature and well-developed subject, and one might wonder what important problems remain to be solved. Both the physical understanding and the numerical calculation of unstable resonator eigenmodes have nonetheless remained rather murky areas; and at the same time detailed knowledge of unstable resonator modes has become increasingly important both for high-power injection-seeded laser design and for evaluation of the excess noise factors discussed in the preceding section. Also, the very useful properties of stable and especially of unstable variable-reflectivity-mirror (VRM) laser activities have been increasingly recognized in recent years. In addition to two partly tutorial

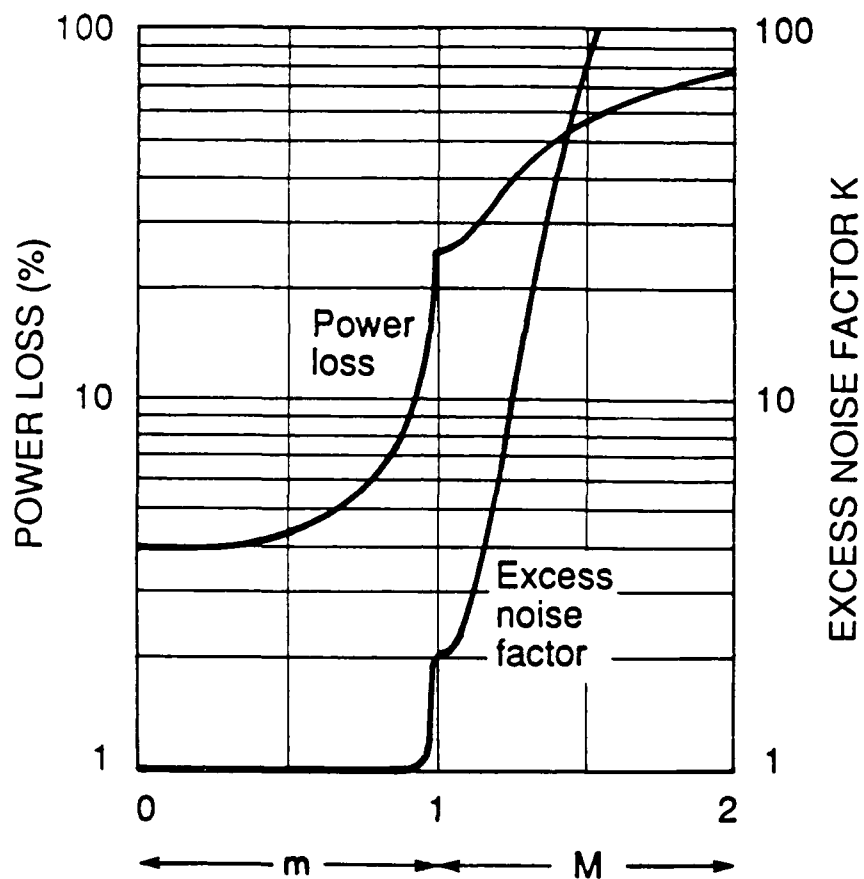


Figure 4. The power loss per round trip (i.e., the effective output coupling) and the excess noise factor K_p for the lowest-order of a laser cavity using a fixed variable-reflectivity mirror, versus the geometric magnification of the cavity optics.

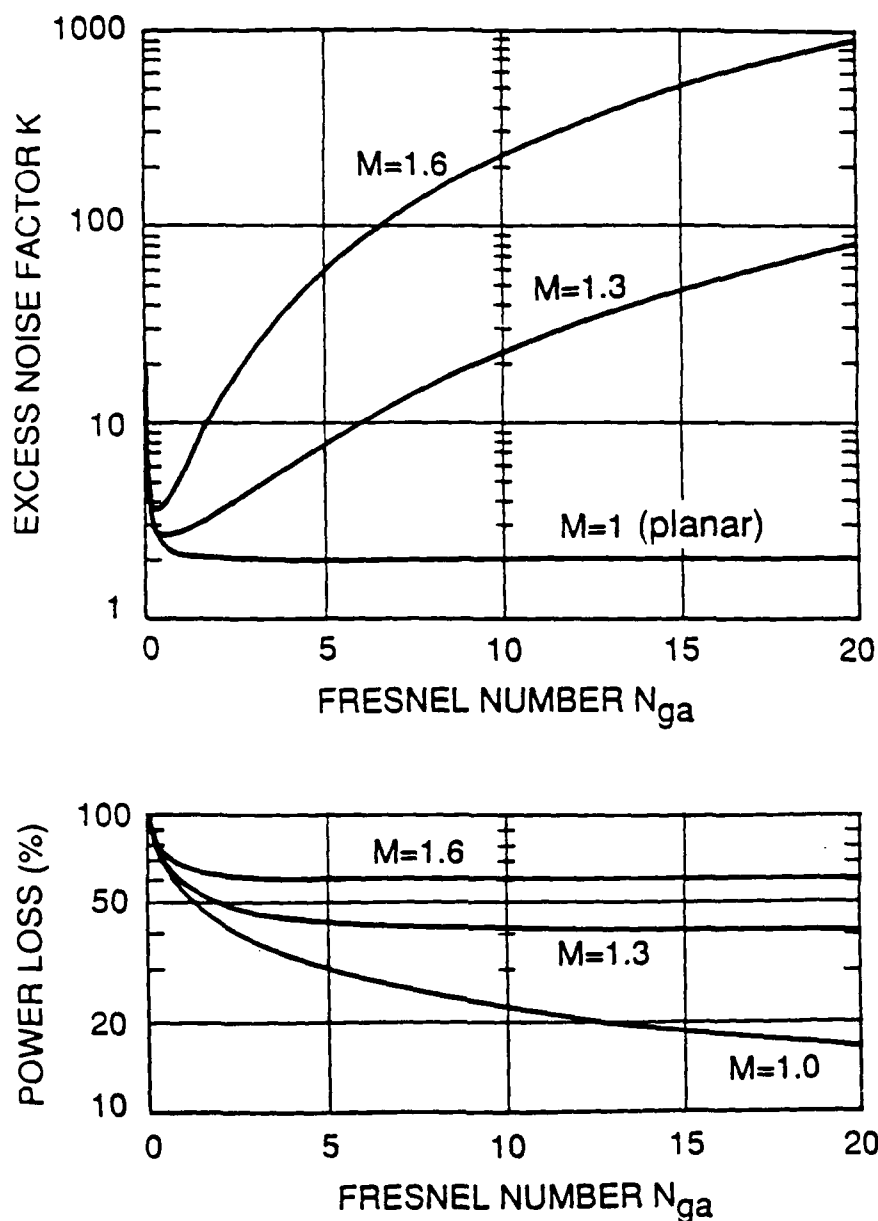


Figure 5. Output coupling and excess noise factor versus mirror aperture or Fresnel number of the variable-reflectivity mirror, for different geometrical magnifications of the cavity optics. The results shown are for the lowest-order cavity mode; higher-order modes have significantly larger values of both quantities.

papers on laser cavity modes, therefore, as published in

A. E. Siegman, "Axial modes in a grating-dispersed laser cavity," *Appl. Phys. B.* **42**, 165-166 (1987).

A. E. Siegman, "Advances in laser resonator design using variable reflectivity mirrors," *OSA Tutorial Series*, in press (July 1988).

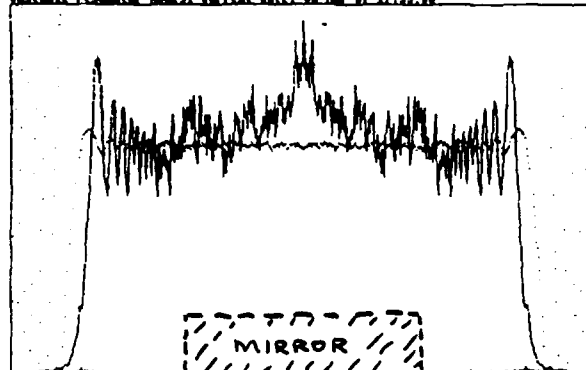
we have during this contract period developed and programmed a very useful reformulation of the "virtual source" approach to unstable resonator theory, following the earlier approaches of Horwitz, Southwell, and other authors. This approach, which will be submitted for publication in the near future under the title of

Paul L. Mussche and A. E. Siegman, "Virtual source theory for unstable resonators: Reformulation and some new results," to be submitted for publication.

now gives us a very clear physical understanding of the fundamental role played by the Fresnel-Sommerfeld edge waves in unstable resonator modes. It also permits us to calculate with high accuracy on a modest personal computer detailed eigenmodes of large-Fresnel-number unstable resonators which would have required a Cray-class supercomputer using earlier methods. Figure 6 shows, for example, the three lowest-order unstable resonator eigenmodes for an unstable resonator with magnification $M = 1.9$ and equivalent Fresnel number $N_{eq} = 49.4$ as calculated more or less instantly on a Macintosh desktop computer. These modes require on the order of 8000 points to plot accurately, yet are calculated using only a dozen edge-wave terms in a virtual source expansion.

These calculations also permit us to explore the exotic biorthogonality properties of unstable resonators, and to obtain much greater understanding, for example, of how optimum external coupling into an unstable cavity requires "adjoint coupling," rather than the conventional "mode-matched coupling," and how optimum coupling into the lowest mode of such a cavity necessarily requires coupling into several higher-order modes as well. Many of these ideas will be summarized in another pending publication,

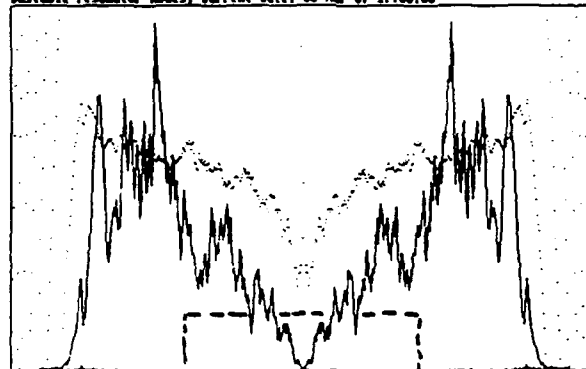
Unstable resonator modes, current date: 83 Mar 87 17:33:00



EVEN
MODE
#1

Effective Fresnel no. = 49.40, $M=1.90$
 $N = 14$, even mode no. 1
 Eigenvalue mod = 1.825369, arg = -3.252974E-03
 $X = -2.5 \rightarrow 2.5$
 Hard copy (1/f) :t

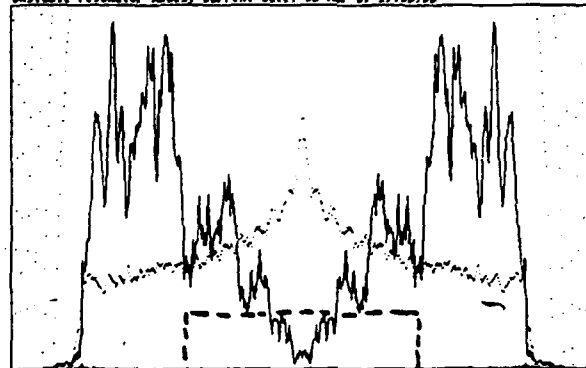
Unstable resonator modes, current date: 83 Mar 87 17:33:00



EVEN
MODE
#2

Effective Fresnel no. = 49.40, $M=1.90$
 $N = 14$, even mode no. 2
 Eigenvalue mod = 0.7609357, arg = -0.3704143
 $X = -2.5 \rightarrow 2.5$
 Hard copy (1/f) :t

Unstable resonator modes, current date: 83 Mar 87 17:33:00



EVEN
MODE
#3

Effective Fresnel no. = 49.40, $M=1.90$
 $N = 14$, even mode no. 3
 Eigenvalue mod = 0.7639345, arg = 0.3636433
 $X = -2.5 \rightarrow 2.5$
 Hard copy (1/f) :t

Figure 6. The three lowest-order even-symmetry modes of a hard-edged strip unstable resonator with $M = 1.90$ and $N_{eq} = 49.4$, as calculated using the virtual source approach. The solid lines are the intensity profiles; the dotted points indicate the phase front deviation. It requires ~8000 points to plot the mode patterns but only ~14 virtual source terms to describe them in the edge-wave expansion.

A. E. Siegman, "The oddball properties of unstable resonators," (Ives Medal Address), J. Opt. Soc. Am. A, to be submitted for publication.

which will be prepared within the next few months also.

VII. MISCELLANEOUS JOURNAL PUBLICATIONS AND REPORTS

In addition to the primary research publications summarized in the preceding sections, we were also called on to prepare a lead review article to help the NAE celebrate the 25th anniversary of the laser, i.e.,

A. E. Siegman, "The Laser—Still Young After Twenty Five Years?" lead article in *Lasers: Invention to Application*, National Academy of Engineering Symposium on the Twenty-Fifth Anniversary of the Laser, ed. by Jesse H. Ausubel and H. Dale Langford (National Academy Press, 1987); pp. 1-16.

Other authors in this volume include Schawlow, Patel, and de Maria.

Another minor publication during the contract period was an outgrowth of the increasing use of personal computers in our research, namely

A. E. Siegman, "Two-dimensional calculations using one-dimensional arrays, or "Life on the Skew," *Computers in Physics*, 74-75 (November/December 1988).

Although written partly "for fun," this note in fact discussed a serious point, namely the utility of an unusual form of "periodically skewed boundary conditions" for handling physical calculations based on two-dimensional arrays with periodic boundary conditions.

VII. PH.D. DISSERTATIONS

The Ph.D. research of Dr. Charles Barker, already described in Section II of this report, was also reported in the Ph.D. dissertation

Ultrafast Nonlinear Optical Measurements using the Tunable-laser-induced Grating Technique, Ph.D. Dissertation, Department of Applied Physics, Stanford University, December 1988.

Dr. Barker is now in the Laser Physics Program at Lawrence Livermore National laboratories.

The Ph.D. research of Adnah Kostenbauder was substantially completed during the reporting period and will be reported in the dissertation

Picosecond Measurements of Dember Potentials Associated with Photocarrier Gratings, Department of Electrical Engineering, September 1989 (est.).

which will be completed in the early stages of our current research program.

VIII. BOOKS

Although preparation of the following reference book and textbook on lasers by Professor Siegman was not directly supported by the AFOSR, this book was completed and published during the current contract period, and its contents owe a great deal to AFOSR-supported research activities over earlier years:

A. E. Siegman *Lasers* (University Science Books, 1986).

The book has been very favorably reviewed in numerous journals, and is believed to be widely used by industrial and research workers in the laser field, as well as by students.

IX. HONORS AND AWARDS

At the 1987 Annual Meeting of the Optical Society of America, Professor Siegman was awarded the Frederic Ives Medal, described as "the highest award of the Society for overall distinction in optics." In April 1988 he was elected a member of the National Academy of Science, and at the April 1989 CLEO/LEOS Meeting in Baltimore, he received the 1988 Quantum Electronics Award from LEOS for outstanding technical contributions to quantum electronics.

Virtually all of Professor Siegman's research during the period leading up to these awards has been supported by the Air Force Office of Scientific Research.

During several years of the reporting period Adnah Kostenbauder was the holder of a Newport Research Foundation fellowship.

Stimulated Wood's Anomalies on Laser-Illuminated Surfaces

ANTHONY E. SIEGMAN, FELLOW, IEEE, AND PHILIPPE M. FAUCHET, MEMBER, IEEE

Abstract—Spontaneous, highly periodic, often permanent surface gratings or "ripples" can develop on the surface of almost any solid or liquid material illuminated by a single laser beam of sufficient intensity, under either pulsed or CW conditions. The grating periods are such that the incident laser beam is diffracted into a tangential wave which skims just along or under the illuminated surface. These spontaneously appearing surface ripples are generated by a runaway growth process analogous to stimulated Brillouin or Raman scattering or small-scale self focusing, but having many of the same properties as Wood's anomalies in diffraction gratings. Hence, it seems appropriate to refer to these spontaneous surface structures as "stimulated Wood's anomalies."

I. INTRODUCTION

A. Experimental Observations

EXPERIMENTAL results extending as far back as 20 years [22], [23] show that when almost any material is illuminated with a single uniform laser beam of sufficient intensity, the material surface may develop spontaneous, highly periodic, and often permanent surface structures or "ripples." Spontaneous surface structures of this type can be produced using single laser beams at many different wavelengths, on a large variety of semiconductor, metal, dielectric, and more exotic materials. Very similar ripples are also observed in a large variety of experiments on laser-assisted film growth [32], laser etching [87], photodeposition [48], [56], and laser materials processing [40]. Ripples of this type often appear on surfaces which can support one or another type of plasmon, polariton, or surface-guided wave [26], [55], and these surface waves then clearly play a significant role in the ripple formation. Similar ripples are also seen with equal ease, however, on other surfaces where no bounded or trapped form of surface wave can be present [31].

For a single laser beam arriving at normal incidence on an opaque surface which does not support any kind of surface wave (e.g., an unmelted semiconductor surface in the visible), the spontaneously generated surface ripples

will normally have a period very close to the incident laser wavelength in the medium above the surface, implying that there is strong self-induced diffraction of the primary laser beam into waves which travel just at grazing angle along the material surface [24].

By contrast, for metals or molten semiconductor surfaces which may support plasma waves, or layered or thin-film surfaces which may support guided surface waves, or for transparent dielectric materials close to the Restrahl region, there may be a significant dependence of the ripple period on the electromagnetic or waveguiding properties of the surface. The surface ripples will generally take on a period such that one of their primary diffraction orders has a k vector component along the surface which is in near synchronism with the relevant surface wave [48]. In any of these cases, if the primary laser beam moves away from normal incidence, the spontaneous surface structures typically split up into complex ripple patterns showing multiple periodicities, with ripples having different periods occurring in either overlapping or separate regions of the surface.

Surface structures of this type can be produced in some cases by a single laser pulse having a duration as short as a few picoseconds [50] if the pulse has an energy density sufficient to produce a significant phase change (e.g., melting or annealing) in the material surface in one shot. Very similar structures may also appear, however, following irradiation of a surface by tens or hundreds of much lower energy pulses at energies well below single-shot damage thresholds [38], [43], [60], [73]. In still other cases, very similar structures can be generated and then erased with a continuous laser beam of sufficient intensity [80], [81]. In all of these cases, the ripples may have a dynamic development both during and after the laser illumination, and in the majority of cases, a permanent grating pattern will remain on the surface after illumination.

B. Theoretical Explanation

According to present understanding, these ripples appear as the result of a nonlinear growth process which is initiated by the scattering of a small amount of light out of the primary laser beam by random irregularities initially present in or on the illuminated surface. These initial noise irregularities may be random variations in surface height, electron density, defect density, or any other optically significant physical property [65], [62]. Inter-

Manuscript received February 2, 1986; revised March 31, 1986. This work was supported in part by the U.S. Air Force Office of Scientific Research. The work of A. E. Siegman was supported by the Alexander von Humboldt Foundation. The work of P. M. Fauchet was supported by IBM at Stanford University, Stanford, CA and Princeton University, Princeton, NJ and by the National Science Foundation at Princeton University.

A. E. Siegman is with the Department of Electrical Engineering and the Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305.

P. M. Fauchet is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

IEEE Log Number 8608949.

ference between any one individual angular component of this scattered light and the primary laser beam then produces a sinusoidal component of spatial variation in the light intensity falling on the material surface, and this sinusoidal intensity variation may cause the corresponding spatial frequency component of the surface irregularity to increase in amplitude.

But each such spatial frequency component in the surface irregularity will then have exactly the correct period to diffract additional light intensity from the primary beam into the corresponding angular component of diffracted light, thus increasing the strength of the surface intensity variation, which in turn can cause still further growth of the corresponding surface irregularities. The net result can be, at sufficient incident laser beam intensity, a very rapid runaway growth, starting from noise, of both the surface irregularities and the corresponding scattered light components, in a fashion closely analogous to stimulated Brillouin or Raman scattering, small-scale self-focusing via the optical Kerr effect, or other stimulated nonlinear optical processes.

From experimental and theoretical results, it is clear that the most rapid growth rates generally occur for a narrow range of spatial frequency components which diffract light into a scattered k vector that is either nearly parallel to and just above the illuminated surface or that is synchronous with any type of polariton or waveguide mode which may be supported by the surface. The spatial component with the highest exponential gain eventually dominates over all other spatial periods, leading to the development of highly regular and periodic ripples or surface structures with periods which match one or the other of the above conditions.

C. Stimulated Wood's Anomalies

As we will point out in this paper, the elementary properties of these spontaneous ripples have a substantial similarity to the familiar Wood's anomalies observed in diffraction gratings and in other periodic surface structures. (This close similarity was first pointed out to us by Zhou Guosheng.) By analogy with the other stimulated nonlinear optical processes mentioned just above, we therefore suggest that these spontaneously appearing surface ripples might well be called "stimulated Wood's anomalies."

We might note that more localized interference patterns showing either circular or in some cases linear structure have also been observed on laser-illuminated surfaces in situations where these patterns have obviously been triggered by single preexisting defects, particles, or cracks on the material surface (e.g., [37], [39]). Although these localized and defect-triggered patterns may have some basic physics in common with the spontaneous ripples discussed in this paper, we will focus our attention in this review on spontaneous patterns extending over enough regular periods to have a definite grating character to them.

The physics involved in these spontaneously induced surface gratings or stimulated Wood's anomalies, both in

their electromagnetic or optical properties and the induced response of the material surface itself, is obviously both complex and interesting. These structures can, moreover, spontaneously develop and then play an important practical role in many situations of commercial interest, including laser annealing and photoetching, laser film deposition, and all types of laser materials processing including cutting, welding, and heat treating. It seems worthwhile, therefore, to present a brief overview of this subject as a part of the present Special Issue on Dynamic Gratings.

D. Organization of This Paper

The organization of this paper is as follows. In Section II, we present a few examples to introduce the elementary properties of these ripples, and in Section III, we note the close similarity between these properties and the properties of Wood's anomalies. In Section IV, we present a general overview of current theoretical approaches to the growth of these ripples and summarize a few basic theoretical results, and in Section V, we summarize briefly some related and more detailed experimental results that have been seen with such ripples. Finally, in Section VI, we comment briefly on some of the practical applications areas where the effects of these widespread ripple phenomena may be important, in either negative or positive fashions, followed by an extensive bibliography of literature citations on spontaneous laser-induced ripple phenomena.

We might note that the literature on both the experimental and theoretical aspects of these single-beam laser-induced surface structures has now become sufficiently extensive and diffuse that it is difficult even to organize it into categories, much less summarize all of it. Hence, while we have attempted to note in the bibliography all the literature citations presently known to us, this paper is intended only as an introductory overview and not as a definitive treatment of the subject of surface ripples.

In addition, while the primary geometrical features and the elementary theoretical aspects of ripple formation seem clear and well understood, the detailed mechanisms involved in the ripple growth process, and especially the exact nature of the material surface response, are very complex and strongly material dependent. Hence, our theoretical discussion in particular should be taken only as an introduction to some of the principal ideas involved, and not as a definitive review of this topic.

II. ELEMENTARY RIPPLE PROPERTIES

In this section, we summarize some of the elementary properties of stimulated Wood's anomalies, primarily under the simplest conditions, e.g., laser intensities which are not too large and opaque surfaces (primarily semiconductors) which do not support any sort of surface wave.

A. Ripple Periods

Fig. 1 is a photograph, for example, of an initially amorphous, As-implanted silicon surface which has been

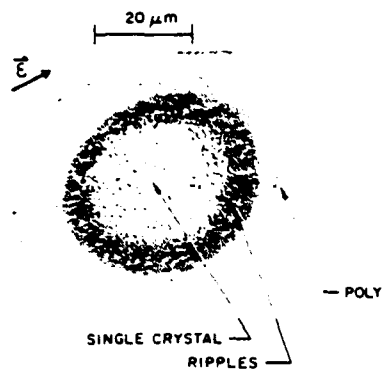


Fig. 1. Photograph of an As-implanted and amorphized silicon surface after irradiation by a single normally incident, linearly polarized, 532 nm laser pulse with a Gaussian transverse intensity profile and a peak intensity of $\approx 300 \text{ mJ/cm}^2$. Direction of the optical electric field vector is indicated by the arrow in the upper left corner. From Fauchet and Siegman [51].

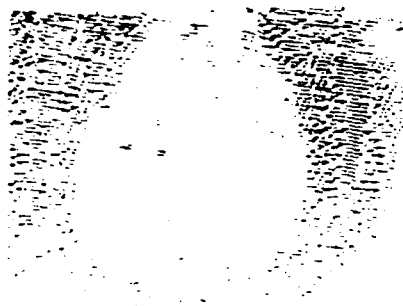


Fig. 2. Germanium surface hit at normal incidence by approximately 100 Q-switched Nd:YAG laser shots with slightly elliptical polarization and intensity close to the single-shot melting threshold. Principle axis of the elliptical polarization is oriented slightly clockwise from the vertical direction in the figure. The dominant ripples have a period equal to the $1.06 \mu\text{m}$ laser wavelength, and a grating k vector parallel to the dominant direction of the incident E field.

hit by a single 80 ps long pulse of linearly polarized 532 nm laser radiation [50]. The normally incident laser beam in this case had a clean Gaussian intensity profile with a peak energy fluence at the center of $\approx 300 \text{ mJ/cm}^2$, which is higher than the melting fluence for amorphized silicon at this wavelength. The shiny central zone approximately $25 \mu\text{m}$ in diameter represents the region within which the laser intensity was sufficiently large to melt and recrystallize the initially amorphous silicon. The sharp outer boundary at approximately $70 \mu\text{m}$ diameter marks the region within which the single-shot intensity was still large enough, if not to melt and anneal, at least to convert the amorphized Si into a darker, presumably polycrystalline form.

In the annular region between these two diameters can be seen a pattern of regular ripples which have been embossed into the silicon surface by the single laser shot. These ripples have a period very close to the incident laser wavelength (in air) and a dominant orientation with the grating vector k_g of the surface grating parallel to the linearly polarized incident E field direction.

Fig. 2 similarly shows a germanium surface which has been irradiated with approximately 100 Q-switched, 100

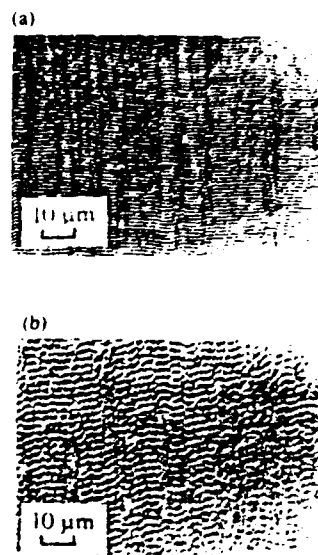


Fig. 3. Germanium surfaces hit at (a) normal incidence and (b) off-normal incidence by a large number of linearly TM-polarized Q-switched Nd:YAG laser shots with intensity substantially below the single-shot melting threshold. The upper photograph, for normal incidence, shows regular fringes with fringe spacing very close to the laser wavelength. The lower photograph, for off-normal incidence, shows intermingled and overlapping areas of coarser and finer fringes. From Young *et al.* [62].

ns duration Nd:YAG laser pulses at normal incidence and with a slightly elliptical degree of polarization. The single-pulse energy fluence in this case was close to, but probably slightly below, the single-pulse melting threshold, although the central portion of the irradiated area has clearly been melted by the repeated pulses. Substantially more regular surface ripples than in Fig. 1 are seen in this case, both in the central melted zone and in the surrounding areas, with ripple spacing again very close to the incident laser wavelength, and with grating vector k_g parallel to the dominant axis of the elliptically polarized beam.

Fig. 3, taken from Young *et al.* [62], shows two other typical photomicrographs of portions of Ge surfaces which have been hit by a large number of repeated, linearly polarized $1.06 \mu\text{m}$ Q-switched laser shots at intensities well below the single-shot damage threshold, at either (a) normal or (b) off-normal incidence. The E field vector points in the vertical direction in both of these figures, and the ripple spacing in Fig. 3(a) is again very close to the incident laser wavelength.

For the case of off-normal incidence, the laser is polarized with the E field in the plane of incidence. (There seem to be opposite conventions in different parts of the optical literature as to whether this should be referred to as s or as p polarization; we will use instead the terms TM and TE polarization, referenced to the plane of incidence.) Fig. 3(b) then shows the typical result for off-normal incidence—that is, the ripples break up into two distinct spacings, one with shorter and one with longer period than in Fig. 3(a), with these two separate periods sometimes coexisting and sometimes occupying separate

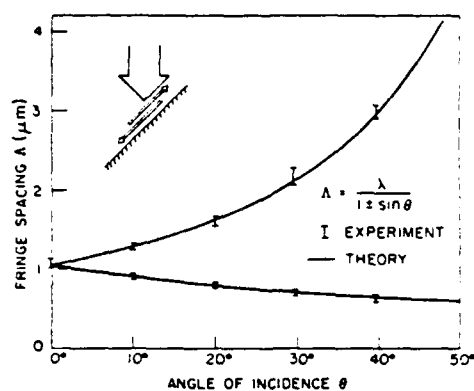


Fig. 4. Measured ripple spacings versus angle of incidence for silicon illuminated by TM-polarized 1.06 μm laser pulses.

areas on the material surface. Fig. 4 shows further data on how the measured spacings for these two periods typically vary with angle incidence, in this case, for 1.06 μm pulses incident on an amorphous silicon surface. These results, along with many others to be found in the literature, illustrate two of the most elementary features of the spontaneous ripple structures on these types of materials, namely the following.

1) With normally incident, linearly polarized light, the ripples form first with the ripple crests and troughs running perpendicular to the E field direction, i.e., the induced grating has a grating vector k_s in the surface plane parallel to the E field direction. (To avoid ambiguous phrases such as "the ripple direction," we will from here on in this paper always characterize ripples in terms of their grating k vector in the surface plane.)

The ripple period Λ produced by a normally incident laser beam, at least on these types of materials, is given to within an accuracy of a few percent by $\Lambda \approx \lambda/n_0$ where n_0 is the refractive index and λ/n_0 is the wavelength in the medium above the surface. Irradiating silicon at normal incidence through water, for example, decreases the ripple period by essentially the refractive index of the water.

2) With off-normal incidence and TM-polarized light, the ripples form with grating vectors k_s in the plane of incidence, and hence parallel to the E field component in the surface plane, and with periods given by

$$\Lambda \approx \frac{\lambda}{n_0[1 \pm \sin \theta]} \quad (1)$$

where θ is the deviation of the laser beam from normal incidence and λ/n_0 is again the wavelength in the medium above the surface. These periods correspond, as shown in the inset to Fig. 4, to those grating spacings which will diffract energy from the incident laser beam at grazing angle either downward along the surface (long-period ripples) or upward along the surface (short-period ripples). They also correspond, as we will note in the following section, to the Rayleigh wavelength condition for Wood's anomalies to occur at the same incident laser wavelength and angle of incidence.

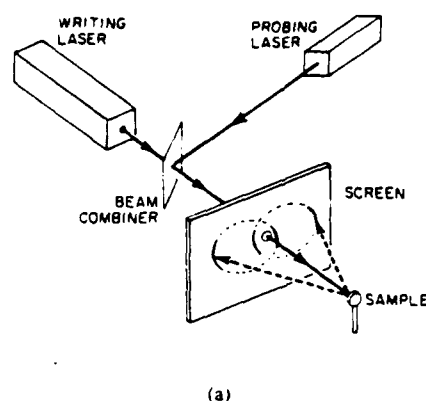


Fig. 5. (a) Back-diffraction technique for observing periodic surface grating structures. (b) Partially filled circular back-diffraction pattern produced by surface ripples having a single constant grating period, but slightly varying grating k vector directions, such as are typically produced by a linearly polarized writing beam at normal incidence. Patterns consisting of two intersecting ovals, such as that sketched on the screen in part (a), will more often be observed for the ripple patterns produced by a laser beam at off-normal incidence.

B. Diffraction Observation of Surface Ripples

Photomicrographs of spontaneous ripples such as those in Figs. 1–3 can be obtained only with some effort since the ripple periods produced by visible laser beams approach the resolution limit of simple optical microscopes. The ripple structures can, of course, be examined in more detail using SEM photographs, thereby revealing more detail about the ripple profiles. Micro-Raman spectroscopy, using a Raman laser source focused into a very small spot which is scanned across the ripple profiles, can also reveal much information about the local variation of materials properties within the ripple peaks and valleys [141], [142].

By far the simplest and most effective method for studying the grating characteristics of these surface structures, however, is to observe the back-diffraction properties of the surface ripples in reflection, using an experimental arrangement such as in Fig. 5(a). This technique, as demonstrated by Young *et al.* [61], [62] and by Haneman and Nemanich [53], requires only a normally incident monitoring laser beam with a wavelength somewhat shorter than the primary writing beam and a means for observing the back-diffracted monitor beam on a suitable screen.

The grating vector components of the surface ripple structures then map directly into geometric patterns on the

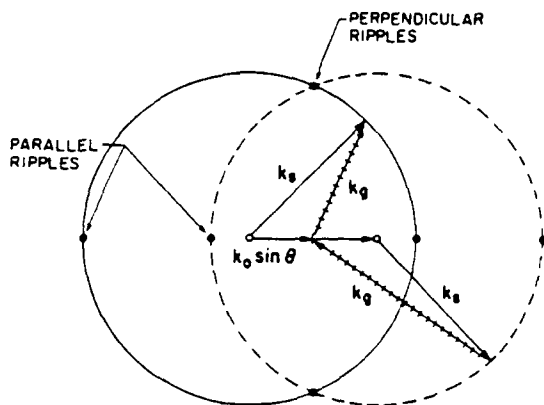


Fig. 6. First-order vector conditions for an incident light beam k_0 , at angle of incidence θ to be scattered by a surface grating with grating vector $\pm k_g$ into a near-grazing scattered wave k_s . The solid circle corresponds to $k_0 \sin \theta + k_g = k_s$, the dashed circle to $k_0 \sin \theta - k_g = k_s$. Note that the magnitude of the scattered wavevector k_s remains constant, independent of its direction.

observing screen, as shown in Fig. 5(a) and (b). Fig. 5(b) shows the partially filled circular pattern which typically results for ripples formed with linearly polarized, normally incident light. This pattern indicates that the grating patterns in different portions of the spot illuminated by the monitoring beam all have the same grating period, but slightly different angular orientations of the grating vector k_g .

When the arbitrarily polarized laser beam is incident on a surface at an angle of incidence θ away from the normal, ripples traveling in additional directions along the surface are often formed, especially at larger intensities or with larger numbers of shots. If $k_0 \sin \theta$ is the k vector component of the incident light in the surface plane, k_g is the grating k vector, and k_s is the k vector of the first-order scattered light, also in the surface plane, then the first-order theoretical (and experimental) synchronism condition for ripple formation is

$$k_0 \sin \theta \pm k_g = k_s \quad (2)$$

as illustrated in Fig. 6. If the grating k_g vector is parallel to the $k_0 \sin \theta$ component in the plane of incidence, and the scattered wave k_s is a grazing wave with the same phase velocity as the incident wave, then (2) leads to the two ripple spacings given in (1). If gratings form at right angles to the plane of incidence, then the two ovals intersect in a single point corresponding to a grating spacing of

$$\Lambda \approx \frac{\lambda}{n_0 \cos \theta} \quad (3)$$

These are so-called "perpendicular ripples" sometimes formed at higher beam intensities.

The lines sketched on the screen in Fig. 5(a) show the general pattern of two intersecting ovals which will be produced by back diffraction from the general family of first-order spontaneous gratings which may be formed by a laser beam at off-normal incidence, as given by the gen-

eral solution to (2) at arbitrary angles. The two innermost arcs represent back diffraction of the monitoring beam by the large-period ripples (producing a small diffraction angle for the monitor beam), while the outermost arcs come from back diffraction by the small-period ripples (large diffraction angle for the monitor beam). Concentration of diffracted light intensity near the centers of these arcs indicates that the ripple k vectors are concentrated along the direction parallel to the plane of incidence, as is customary for linearly TM-polarized irradiation, at least at lower intensities. If spontaneous gratings are also generated with grating k vectors oriented in other directions along the surface, as occurs for elliptical or circularly polarized light at higher intensities, then the complete intersecting-oval pattern of back-diffracted light will be filled in. The grating spacing for perpendicular ripples given in (3) comes, of course, at the intersection of the two general ovals.

C. Ripple Polarization Properties

Spontaneous surface ripples are by far the most readily obtained using linear TM-polarized laser beams, and the resulting ripples then often have grating k vectors which are closely concentrated around the E field direction in the surface plane. Ripples formed with linearly TM-polarized light under single-shot or high-intensity conditions, especially where there is considerable macroscopic damage to the surface, may exhibit a spread of ripple k vector directions distributed over 5 or 10° about the E field direction. Ripples formed under more controlled conditions, e.g., ripples which build up slowly over a large number of low-intensity shots tend to exhibit much more regular and well-aligned fringe patterns with a grating k vector confined to a narrow range of angles very accurately parallel to the E field and the plane of incidence.

The use of circularly polarized light at normal incidence often strongly inhibits the formation of ripples. If the laser pulse energy or the number of repeated pulses is increased, ripples will eventually form with circularly polarized light, with the same period equal to the laser wavelength as for linearly polarized light, but with grating k vectors oriented in all directions along the surface. The circular back-diffraction pattern shown in Fig. 5(b) will then be uniformly filled in around the circular arc. The use of elliptically polarized light at normal incidence will typically fill in this pattern partially around the circle.

The polarization properties become more complex for off-normal incidence, with considerable variation among different materials and conditions of illumination. At angles of incidence only slightly off normal, the use of elliptical, circular, or unpolarized light may lead at first to the formation of ripples parallel to the plane of incidence, but with two slightly different periods, as shown in Fig. 3. Continued illumination or the use of higher intensities may then generate additional ripples traveling in other directions on the surface, so that the pattern of intersecting ovals shown in Fig. 5(a) becomes partially or completely filled in. There is some evidence that with circularly or elliptically polarized light, this intensity pattern will be

filled in assymmetrically, i.e., with a handedness that matches the direction of rotation of the light polarization [61], [62].

For larger angles of incidence, on the other hand, with almost any type of polarization, the ripple patterns will often show a rectangular symmetry relative to the plane of incidence. One or both of the two primary grating k vectors parallel to the plane of incidence may develop first. Then, after further pulsed illumination, strong perpendicular ripples with grating k vector perpendicular to the plane of incidence, as given by (3), may emerge. In some case, only these perpendicular ripples are seen.

Even in the simple case of normally incident, linearly polarized radiation, where the strong parallel ripples such as shown in Fig. 2 develop first, if one keeps hitting the same surface with repeated pulses, eventually strong perpendicular ripples may also develop on top of the parallel ripples, giving the surface structure a rectangular symmetry indexed to the E field direction of the incident light [73].

We might note again that photomicrographs of surfaces which have developed complex overlaid grating patterns are relatively uninformative as to just which ripple periods and directions are actually present. The back-diffraction pattern, very much like an atomic X-ray pattern, gives much more direct indication of which periodic structures are present.

D. Ripple Growth and Propagation

Depending on the laser wavelength and the nature of the illuminated material, the spontaneous ripples or stimulated Wood's anomalies we are discussing here may build up under at least three broadly different classes of conditions.

1) Ripples are readily formed on many materials by single short-duration laser pulses having an energy fluence sufficient to produce some form of phase transformation (melting, vaporization, partial, or complete annealing) in the laser surface during a single laser shot. There are many examples of such pulsed laser-induced surface transformations; nearly all of them seem to produce ripples. Fully formed ripples have been observed on semiconductor and metal surfaces following illumination by single laser pulses as short as 7 ps in duration [100], [101].

2) Ripples are also formed on certain materials after multiple pulse illumination with tens, hundreds, or even larger numbers of pulses, at low repetition rates and at energy fluences substantially below those needed to produce any easily observable single-shot effects. Common examples of this type of behavior include germanium surface illuminated with low-energy visible or near-infrared pulses [73], silicon surfaces, especially amorphous or thin-film silicon [38], [43], [73], and some metals when illuminated in the same spectral region. There appear to be other classes of materials, however, for which this multiple-pulse ripple formation does not occur, for example, with 10.6 μm pulses on quartz or glass where the

primary ripple growth mechanism is believed to be direct vaporization of surface material [136], [137].

3) Finally, ripples can be produced by continuous (CW) laser illumination, either at an intensity sufficient to produce melting or other surface transformation, as in CW cutting, welding, or laser annealing [80], [81] or at much lower intensities under conditions where some cumulative surface modification can integrate with time, for example, laser etching or laser photodeposition experiments [87], [88], [56]. We will mention further examples of these cases below.

Under all of these conditions, if the laser beam is moved transversely with time or between successive shots in the direction of the ripple k vector, the ripples in the new beam location will build up in spatial phase with those in the previous beam location, evidently triggered by diffraction from the previously created ripples in the overlap region between new and old locations. Using this technique, it is possible to "propagate" coherently phased ripple patterns over large distances in the direction of the ripple k vector.

If the laser pulse is translated between shots in a direction perpendicular to the grating k vector, on the other hand, there will be in most cases little or no phase coherence between older ripples and newly formed ripples in the new beam location, even if the two laser spots are partially overlapping.

E. Ripples on Other Materials

As we have mentioned in the Introduction, ripples with properties very similar to those described above have been formed using single or multiple laser pulses on a wide variety of semiconductors, including silicon, germanium, gallium arsenide, and indium antimonide, in single-crystal, molten, thin-film, or a variety of amorphous forms. Very similar ripples have also been observed using single laser beams on numerous metal surfaces including aluminum, copper, brass, iron, molybdenum, silver, Ni_3P_1 , films, and liquid mercury. Wavelengths ranging from the ultraviolet to as long as 10 μm have been used in these experiments. Pulsed CO_2 laser beams at 10.6 μm have also been used to write similar gratings on dielectrics such as quartz, NaCl, and various glasses.

Another broad class of experiments where spontaneous ripples are often seen are laser-induced photoetching, photodeposition, and other photolytic experiments [48], [56], [112], [126], [138]. In many of these cases, the surface etching, film deposition, or other surface modification process has a more or less linear dependence on laser intensity, and hence the surface modification can proceed, albeit slowly, even at low light intensities. Ripple formation can thus be expected in these cases, under proper conditions, even with low-intensity CW light beams.

Earlier surveys of ripple formation have been given by Zhou *et al.* [65], and the ubiquitous nature of this phenomenon has been noted by van Driel *et al.* [60]. In addition to the materials mentioned here, one might also look for spontaneous ripple formation in still other situations

where laser beams modify surface properties—for example, when a high-power laser beam is directed onto the surface of a highly absorbing saturable dye solution. To our knowledge, no experiments have been tried or results reported for this particular case.

III. WOOD'S ANOMALIES

We can next consider the very close similarity between these spontaneous surface ripples and the Wood's anomalies that have long been observed on diffraction gratings.

A. Early Observation

As early as 1902, R. W. Wood noted that on many different kinds of diffraction gratings, the light diffracted into one order would exhibit a sharp dip in diffraction efficiency whenever conditions were such that some other order at the same wavelength was being diffracted tangentially, i.e., at grazing angle along the grating surface. Similar observations were made by many others in subsequent years [1]–[3], [11], [6], [7], [8]–[10], [12].

Fig. 7, taken from a 1921 paper by Ingersoll, shows, for example, in part (a) the light intensity diffracted into first and second orders, plotted versus observation angle (or equivalently versus wavelength) from a speculum grating with a grating period $\Lambda = 2.916 \mu\text{m}$ when this grating is illuminated by white light at normal incidence. The sharp narrow reduction in the first-order diffracted intensity at angles near 30° corresponds (within an error of a percent or two) to the incident radiation at wavelength $\lambda \approx 1.46 \mu\text{m}$ which is being diffracted into first order at these angles. This same wavelength is, however, simultaneously being diffracted into second order at diffraction angles $\pm 90^\circ$ from the normal, i.e., tangential to the grating surface. Note that these dips occur in the perpendicular polarization only.

Weaker dips in the first-order diffracted intensity near 20° and 15° correspond similarly to incident radiation at 0.972 and $0.729 \mu\text{m}$. These wavelengths are diffracted into these two angles in first order, and are simultaneously diffracted tangentially along the grating surface in third and fourth order, respectively. The same $0.972 \mu\text{m}$ radiation which is diffracted at 20° in first order and tangentially in third order also exhibits a sharp dip in its second-order diffraction efficiency at an angle near 42° . Note that all of these dips occur in the perpendicular polarization only.

Part (b) of Fig. 7 is an expanded view of the first-order diffracted intensity in the vicinity of 30° diffraction angle for the case when the primary white-light radiation is incident on the grating not at normal incidence, but at approximately 2.5° away from the normal. The single dip observed for normally incident radiation now splits into two distinct dips, corresponding to two slightly different wavelengths. These turn out to be the two wavelengths which become separately tangential to the grating surface in second order for a primary beam incident at 2.5° , with the second-order radiation in one case going to the right along the grating surface and in the other case to the left.

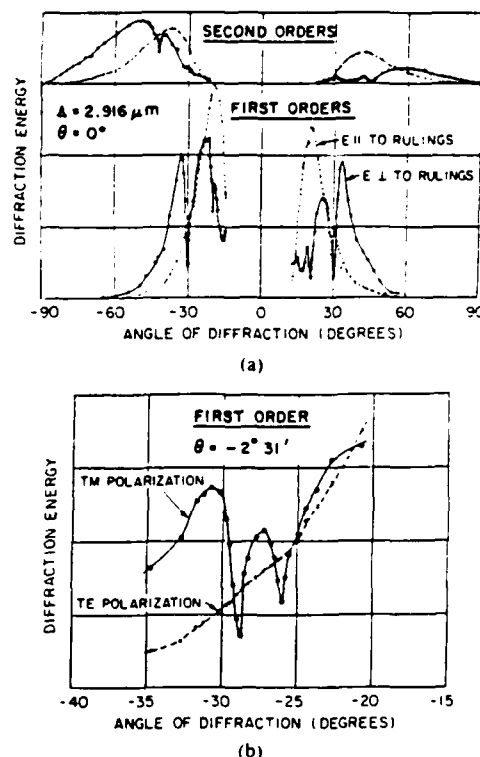


Fig. 7. (a) Measured energy scattered into the first and second orders of diffraction, plotted versus diffraction angle for two different polarizations of the incident light, using a speculum diffraction grating with 8709 lines/in illuminated with white light at normal incidence. The first and second orders are separated prior to detection using an external prism following the grating. Each of the sharp dips in the solid curves corresponds to a different Wood's anomaly. (b) The Wood's anomaly located near -30° in part (a) splits into two separate dips when the angle of incidence of the primary beam is changed to 2.5° off normal incidence. Redrawn from Ingersoll [7].

These two dips thus represent essentially the same phenomena as the two ripples shown in Fig. 4, except that here there is one grating spacing and two wavelengths, rather than one wavelength and two grating spacings.

B. The Rayleigh Diffraction Condition

The general condition for a beam incident on a reflection grating at angle θ from the normal to be diffracted into angle θ' in the m th diffraction order can be written as

$$k_0(\sin \theta + \sin \theta') = \pm mk_g \quad (4)$$

where $k_0 \equiv 2\pi n_0/\lambda$ with λ being the vacuum wavelength of the incident radiation, n_0 is the index of refraction of the medium in which the grating may be imbedded, $k_g \equiv 2\pi/\Lambda$ is the k vector of the grating of period Λ , and m is an integer. The relationship among wavelength, angle of incidence, and grating spacing necessary for the m th order to be diffracted at grazing angle along the surface (i.e., $\theta' = \pm 90^\circ$) is thus given by

$$m\lambda_m = n_0\Lambda(1 \pm \sin \theta) \quad (5)$$

which for $m = 1$ is the same as (1). The wavelengths λ , which satisfy (5) exactly for a given angle of incidence

and grating spacing are often referred to as the *Rayleigh wavelengths*, from the early grating diffraction theories developed by Rayleigh [4], [5].

The wavelengths at which Wood's anomalies are observed are normally given, to within a few percent, by the tangential or Rayleigh diffraction condition of (5). Most experimental observations do show, however, small but real deviations, typically on the order of a few percent or less, between the positions of the observed dips and the exact Rayleigh condition, just as do the spontaneous ripples. These discrepancies can be explained in both cases in part by dispersive effects of the grating material itself, and in part by leaky wave propagation effects on the periodic surface, especially for larger groove depths.

As both parts of Fig. 7 show, strong Wood's anomalies are usually observed only for light polarized in the plane of incidence and perpendicular to the gratings, i.e., for TM-polarized light, again in agreement with the stimulated ripple observations. The Wood's anomalies shown in Fig. 7 consist entirely of strong dips in the diffraction intensity at the Rayleigh wavelengths. Wood's anomalies in the form of intensity peaks occurring at the same wavelengths are also sometimes seen, but only with strongly blazed gratings or gratings having much deeper groove depths.

By making more careful measurements using photoelectric detection, Palmer [12] was also able to show that gratings having larger groove depths (comparable to or greater than a wavelength in depth) could produce weaker anomalies in the TE polarization as well. These TE anomalies occurred at exactly the same angles or wavelengths as the TM anomalies, but with different shapes, more often showing sharp discontinuities in intensity versus angle rather than dips or peaks.

Palmer further demonstrated that immersing a grating in water shifted the locations of the Wood's anomalies to exactly the locations expected using the Rayleigh diffraction condition, but taking into account the index of refraction n_0 of the water. Placing a glass plate on top of the grating but separated from it by several optical wavelengths, on the other hand, did not shift the location of the anomalies, presumably because the wave scattered at grazing incidence skimmed the surface so tightly that it did not see the dielectric effects of the glass plate. An evaporated dielectric film of variable thickness did shift the anomalies in accordance with its dielectric constant, but only when the film thickness was increased to an optical wavelength or more. These results are all consistent with the picture of the anomalies as resulting from a diffracted wave traveling tangentially or at a grazing angle very close to and just outside the (opaque) grating surface.

C. Magic Angles

Palmer also noted that there were certain special angles of incidence θ given by the condition

$$\sin \theta = \frac{m_1 - m_2}{m_1 + m_2} \quad (6)$$

for which the same wavelength λ could meet the Rayleigh diffraction conditions $m_1\lambda = (1 + \sin \theta)\Lambda$ and $m_2\lambda = (1 - \sin \theta)\Lambda$ simultaneously in opposite directions for two different orders m_1 and m_2 . That is, these conditions correspond to the incident radiation being simultaneously diffracted at grazing angles in both directions along the surface. Values of these magic angles include $\theta = 19.47^\circ$ for $m_1 = 2, m_2 = 1$; $\theta = 30^\circ$ for $m_1 = 3, m_2 = 1$; and $\theta = 11.5^\circ$ for $m_1 = 3, m_2 = 2$. Radiation of the proper wavelength might then be expected to exhibit some form of double Wood's anomaly in its diffraction behavior into other diffraction orders. Stewart and Galloway [17] studied some of these crossing points in more detail, finding that the two separate Wood's anomalies which converged at these magic angles seemed to repel each other, rather than enhancing each other.

D. Recent Developments

More extensive discussions of these early experimental observations and of the early theoretical attempts to explain Wood's anomalies may be found in reviews by Twersky [13], [14], and by Millar [15]. Developments in modern guided-wave and antenna technology beginning in the late 1930's also led to renewed interest in Wood's anomalies at microwave as well as optical wavelengths because of their relevance to corrugated surfaces, leaky-wave antennas, and other types of open periodic structures [21]. This led to a series of more modern and extensive theoretical analyses of Wood's anomalies [18], [19] which took much more detailed account of surface impedances, groove depths and shapes, and field polarization effects. These more detailed analyses give much insight into various types of physically similar but mathematically different versions of Wood's anomalies, including "Rayleigh," "leaky-wave resonance," and "Bragg-angle" types of anomalies. They make clear, in particular, that some of these anomalies may occur in both TM and TE polarizations, but only for gratings with larger groove depths. All of the linear aspects of Wood's anomalies seem, in fact, to be equally present and relevant for different examples of the stimulated Wood's anomalies reviewed in this paper.

IV. THEORY

A detailed and general theoretical description of the spontaneous ripple formation process becomes very complex, especially if quantitative predictions are desired. This occurs in part because the electromagnetic theory involved is moderately complicated, but even more so because the surface response mechanisms which close the feedback loop are complex, different in character in different materials, and often highly nonlinear in their response to optical intensity. In this section, therefore we can only give a quantitative outline of the general approach which must be followed to model the stimulated Wood's anomalies.

A. Theoretical Approaches

There are three related but different approaches to the analysis of the ripple growth mechanism which might be pursued, as follows.

1) In what seems to be the most general approach, and the one which we will follow [65], [86], one begins by assuming small and more or less random initial transverse variations in the physical or electromagnetic properties of the illuminated surface. These transverse variations are then separated analytically into individual spatial frequency components or sinusoidal surface gratings along the material surface.

From grating diffraction theory, one can then compute the amount of light which any one of these spatial frequency components will diffract out of the primary laser beam into the appropriate diffraction orders, both above and below the surface. The resulting total light intensity impinging on the surface can then be evaluated from superposition of the primary wave plus these scattered light waves.

Using a suitable model for the physical properties of the material surface, one can then attempt to calculate the rate of growth (or decay) of the initial surface variation which will be produced by the total light intensity, acting through surface heating, thermal expansion, vaporization, photodeposition, or whatever seems to be the dominant physical mechanism. An increased amplitude for the relevant spatial component of the surface variation will then lead to increased light diffraction, which will lead to increased modulation of the incident light intensity, which will lead to increased rate of ripple growth, and so on.

The net result will be a theoretical prediction for ripple growth rate versus laser intensity and versus the period or spatial frequency of the assumed initial surface variation. One can then assume that the spatial frequency component with the highest growth rate will build up faster than any other spatial component, thus determining the period of the ripples that will eventually become dominant. We will say more in detail about this analysis below.

2) A second and closely-related approach, applicable to surfaces which can propagate any sort of plasmon, polariton, or other guided surface wave is to consider the coupling between the incident laser beam and the propagating surface wave caused by the same random initial disturbances in the surface properties [48]. Applying the laser beam can then excite the traveling surface wave in resonant fashion, and this in turn can lead to a spatially varying total light intensity in and along the material surface. This intensity variation may then produce a growth rate for the surface irregularities, increasing the coupling from the laser beam into the surface wave, and leading to the same exponential growth as above.

This approach focuses more attention on the surface wave in those cases where it exists, and thus gives more emphasis to the propagation vector of this wave. As we have seen, however, stimulated Wood's anomalies readily occur even on surfaces which do not propagate any surface waves. The surface-grating or spatial-frequency

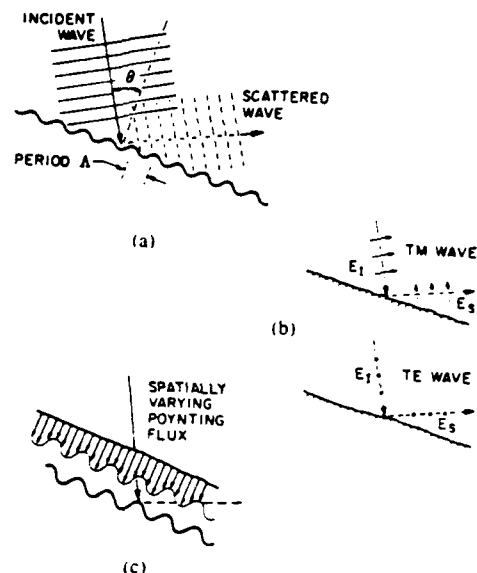


Fig. 8. (a) Scattering of an incident plane wave into a specific direction by one spatial frequency component of surface irregularity having period Λ . (b) Intensity interference between the initial and scattered waves for TM- and TE-polarized beams. (c) Periodic spatial variation of the Poynting vector flux flowing into the sinusoidally modulated surface due to the linear combination of incident and scattered waves. The Poynting vector flux has the same periodic variation as the surface ripples, but may be shifted transversely in phase from them.

approach described above seems more general, therefore, and if properly done, should in the end include the same results as the surface wave approach in cases where these exist, as already pointed out in [48] and [65].

3) Finally, a third approach is to view the assumed surface irregularities as being in the form of individual discrete or point irregularities, and to focus on the net scattering of the incident laser beam by any one such irregularity or on the induced polarization and the resulting fields radiated by this discrete element [41], [125]. One can then add up the scattered fields from multiple irregularities, perhaps assuming these to be arranged in periodic arrays, and look for those conditions where the total field, incident plus scattered, at any one irregularity is most enhanced by the presence of the other irregularities.

This approach is, in essence, a Green's function or delta function version of the spatial frequency approach outlined above. While it may give additional insight in those cases where the ripples form as very narrow stripes or lines (e.g., certain kinds of photodeposition experiments), again we suggest that the spatial frequency approach is probably still best matched to the intrinsically periodic nature of the spontaneous ripples.

B. The Surface Diffraction Problem

In the spatial frequency approach, therefore, the first step is to assume an initially small sinusoidal variation in surface properties [Fig. 8(a)]. Lacking advance information as to the most significant quantities, it may be necessary to consider sinusoidal spatial variations in surface height, temperature, electron density, composition, crystal structure, defect density, surface deposits, or any other

physical quantity which may change the dielectric properties of the material surface (cf., e.g., Wilson and Houle [138]). In our calculations [65], we have typically found that surface height variations produce stronger diffraction effects than other quantities, but other diffraction mechanisms cannot be ruled out. Van Driel and co-workers [61], [62] have introduced the concept of "selvedge" to describe the initial randomly rough surface layer on typical ripple-forming materials.

Given any sinusoidal variation in surface properties, the diffraction of the incident laser beam into relevant diffraction orders above and below the surface can be calculated from standard diffraction theory. It might be noted in this connection that, despite a long history of the subject, accurate calculations of diffracted intensity from a sinusoidally modulated surface, taking full account of the polarization properties of the incident wave and the dielectric properties of the substrate material, are still relatively complex, even for small depths of modulation (cf. eg., Hessel and Oliner [18] and Petit [20].)

When the incident plus diffracted waves are taken into account, the total light intensity impinging on the surface will then automatically have a transverse variation with the same spatial periodicity Λ as the spatial frequency component that produced the diffracted light. Note, however, that a simple calculation of the light intensity $|E|^2$ just above the surface will clearly not account for the mechanism of ripple growth since the TM waves which produce the strongest ripple growth have nearly zero interference between incident and scattered E fields [Fig. 8(b)], while the situation is exactly opposite for the TE case which has the slowest ripple growth. One must evidently calculate instead something like the normally directed Poynting vector in the plane just above the surface, as in Fig. 8(c).

As an example of this type of calculation, Fig. 9(a) shows results we have presented earlier [65] giving the peak-to-peak percentage ripple in the sinusoidally modulated Poynting vector flux, plotted versus the period Λ of the assumed surface ripples, for a normally incident, TM-polarized laser beam of fixed wavelength λ striking a sinusoidally corrugated surface having the dielectric properties of molten silicon. The essential point in this result is that close to the Rayleigh condition $\Lambda = \lambda$, which produces grazing-angle scattering, a very small peak-to-peak surface ripple $h/\lambda = 1$ percent produces a much larger periodic modulation with a peak-to-peak variation of $\approx \pm 15$ percent in the Poynting vector flux flowing into the surface. Generally similar results are also obtained for molten metals or, with somewhat reduced amplitude, for unmelted metals or semiconductors.

For comparison, Fig. 9(b) shows results of the same calculation for a TE-polarized, normally incident wave. The magnitude of the periodic interference effect is reduced by roughly two orders of magnitude compared to the TM case, consistent with the observations on relative spontaneous ripple growth rate with these two polarizations. In addition, rather than a peak (or two peaks), there

is a sharp discontinuity at the Rayleigh condition, which is clearly related to the perpendicular Wood's anomalies mentioned in Section III and studied by Palmer [12]. Results very similar to these can be calculated with many other types of surfaces.

C. Surface Response Mechanisms

One must then complete the feedback loop from initial surface variation to diffracted light and back to increasing surface variation by asking how the periodic spatial variation of the light flux into a material surface may cause the initial modulation of the surface to grow. This is a complex issue for which the answers are, in many cases, unclear since it appears that many different physical mechanisms could potentially be active in stimulated ripple growth on different surfaces. Mechanisms to be considered include

- effects of thermal deposition, rapid temperature increase, and thermal expansion,
- phase transitions such as annealing, melting, recrystallization, and vaporization,
- changes in electron and hole densities,
- various kinds of defect formation,
- saturable absorption,
- plasma formation, both in and above the surface,
- surface chemical changes,
- photoetching
- oxidation, film formation, laser-assisted chemical vapor deposition,

and possibly others. Each of these mechanisms will depend in a complicated way on the local light intensity, and will produce complicated effects both on the surface profile and on the local electromagnetic properties of the surface material. Which of these mechanisms plays the dominant role in many ripple experiments is still far from clear.

D. Feedback Phase

Ripple growth requires, of course, positive feedback in which the diffracted light pattern produced by an initial ripple pattern causes that same ripple pattern to grow rather than decay in amplitude. Satisfying this condition does not seem to be a problem, however, at least for a normally incident laser beam.

Consider, for example, the corrugated surface case assumed in the diffraction calculations of Fig. 9(a). This curve shows, in fact, two closely spaced resonances of similar magnitude but opposite sign, both located very close to the Rayleigh condition. The consequence is that at one of these resonances, the maximum of the Poynting flux coincides spatially with the peaks of the surface corrugations, while at the other, the maximum flux flows into the troughs of the surface corrugation. Growth is thus possible at either one or the other of these closely spaced resonances, depending on whether larger energy flux will cause the corrugation peaks to grow (due to thermal expansion or material deposition, for example) or the val-

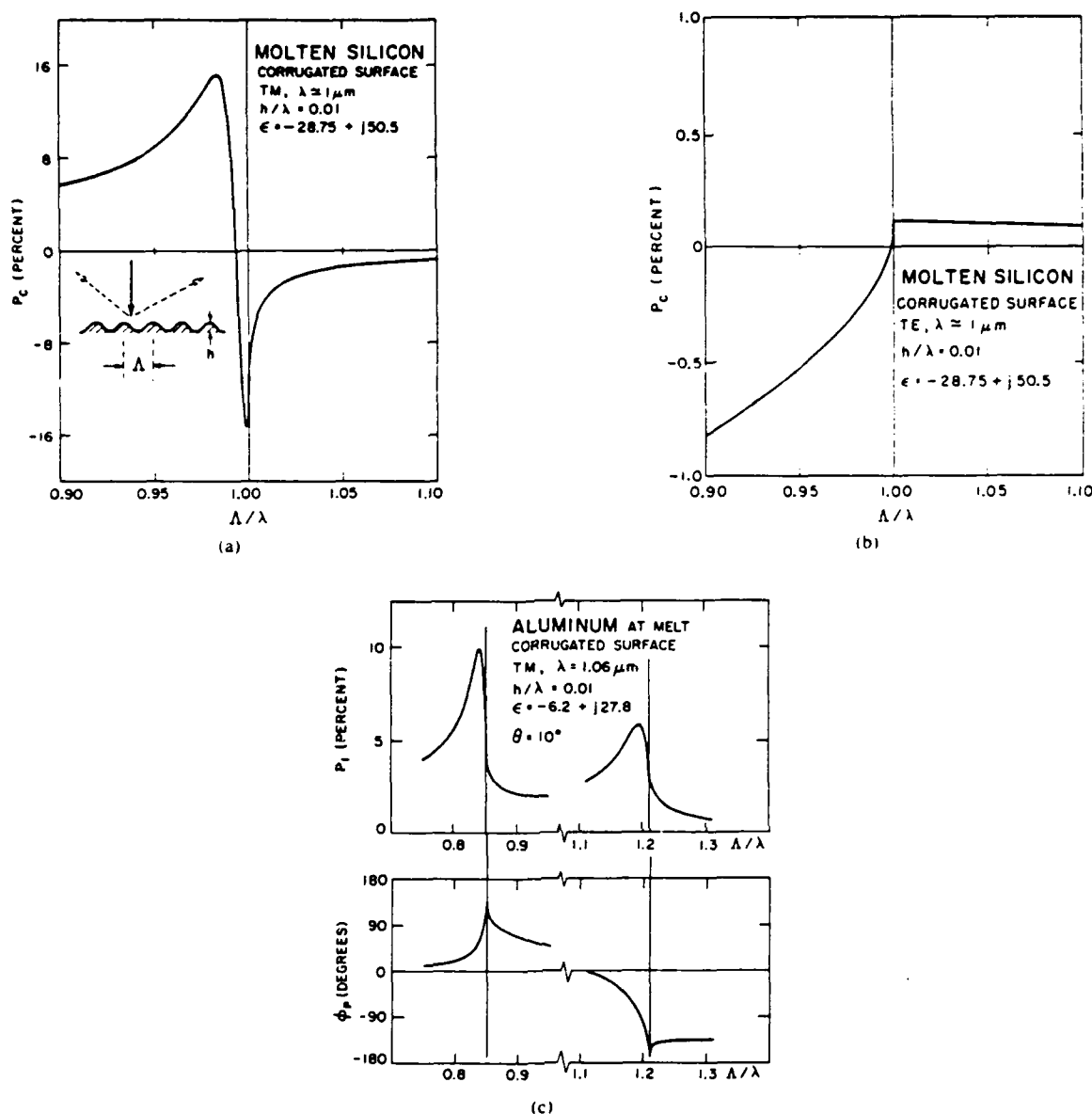


Fig. 9. Three plots of the peak-to-peak amplitude of the sinusoidal spatial variation in the Poynting vector flux (relative to the constant part of the Poynting flux), plotted versus grating period Λ divided by optical wavelength λ , for a surface with 1 percent sinusoidal height corrugations. (a) Molten silicon surface, normal incidence, TM polarization (geometry as shown in the inset). (b) Molten silicon surface, normal incidence, TE polarization. (c) Molten aluminum surface, TM polarization, primary beam 10° off normal incidence.

leys to deepen (due to evaporation, for example). Growth can occur with a surface response of either sign, the only difference being a slight deviation (on the order of a few percent) from exact Rayleigh condition.

In the case of ripples formed by photodeposited film growth, for example, the deposition rate is a function of the strength of the local electric field and the feedback comes from its modulation. Hence, as the initial perturbation grows, the film corrugation deepens and positive feedback is established.

In the case of amorphous silicon illuminated at near the

melting threshold, ripples consisting of alternating stripes of recrystallized and still amorphous silicon are commonly observed. In addition, the surface is found to be corrugated. It appears that a small but positive feedback exists in the solid phase coming initially from the surface roughness, and that when the surface temperature of the sample comes close to melting point, the regions which have absorbed more energy undergo melting first. Since the optical properties of silicon are very strongly modulated when the surface melts, the resulting diffraction and modulation of the Poynting vector flux becomes ex-

tremely large, effectively shutting off energy flow in the not yet molten regions.

E. Phase Shifts at Off-Normal Incidence

Fig. 9(c) shows one further plot of Poynting vector modulation versus periodic spacing, this time for a TM-polarized wave on a weakly corrugated molten aluminum surface at 10° off-normal incidence. There are evidently two sizable resonances in the Poynting vector modulation occurring for grating periods close to the Rayleigh conditions $\Lambda = \lambda/(1 \pm \sin \theta)$, which are indicated by the vertical lines in the plot.

For off-normal incidence, however, the spatial peaks of the Poynting vector flux do not coincide in phase with either the peaks or valleys of the original surface corrugation, but have a phase variation which changes rapidly with grating period, as indicated by the phase angle ϕ_0 in the lower part of Fig. 9(c). This relative phase shift between original surface variation and resulting light flux might lead, in some cases, to a transverse sliding motion as well as growth in time of the surface ripples.

F. Large-Signal and Saturation Effects

The ripple formation process has been discussed thus far largely in terms of a linear feedback mechanism, leading to exponential growth starting from random initial disturbances. We should note, however, that many of the physical mechanisms involved in ripple formation (e.g., melting or vaporization) are highly nonlinear effects, often having discontinuous threshold characteristics. As a result, the overall feedback loop involved in the ripple formation process is likely to be highly nonlinear, and the actual growth to be much more like a discontinuous threshold process than simple exponential growth.

In addition, the diffracted signals from a periodically modulated surface, even with purely sinusoidal modulation, will begin to have a strongly nonlinear amplitude dependence on the depth of the surface modulation as soon as the relative strength of the surface variation exceeds more than a few percent. As the modulation depth becomes larger, the relative strength of the Wood's anomalies for TE polarization or for transverse anomalies also begins to increase. This is in at least qualitative accord with the experimental observations that TM ripples generally grow first on fresh surfaces, but that ripples with perpendicular or TE character will also often appear later on, with larger laser intensities or increased numbers of pulses applied to the same surface area.

Finally, it should also be noted that in most ripple experiments, what one actually observes are the residual ripples left frozen into a surface well after the stimulating laser pulse has ended. One must consider not only how the surface ripples may be formed by rapid processes—sudden melting, vaporization, plasma formation—which take place during the laser pulse, but also, and equally important, how the ripples will be reshaped and transformed by relatively slow processes—surface tension,

mass transport, condensation, recrystallization—that will take place well after the light flux has ended.

V. ADDITIONAL EXPERIMENTAL RESULTS

To give some further insight into ripple properties, we can also note briefly a number of additional interesting effects observed in ripple experiments reported to date.

A. Dispersive Ripple Spacings

We have pointed out that the simple Rayleigh grazing-angle condition for both linear and stimulated Wood's anomalies will be modified somewhat by the dielectric properties of the material surface. The synchronization condition along the material surface can also be modified if the surface develops a dielectric film which modifies the surface propagation velocity and perhaps even traps a guided wave or if the surface electromagnetic properties are such that the surface can propagate either a trapped or leaky wave, such as a plasmon or polariton. Any of these dielectric or dispersive effects will then modify both the wavelength or grating period for linear Wood's anomalies on passive gratings and the dominant growth period for stimulated Wood's anomalies. We mention here a few of the cases where such dispersive effects have been clearly seen in stimulated ripple growth experiments.

It is well known, for example, that silver and similar metal surfaces can propagate strong surface plasmon/polariton modes, especially in the ultraviolet region. Strong dispersive effects, consistent with plasmon wave theory, have thus been seen for ripples produced on metals illuminated with VUV excimer laser pulses [49], [71] or during the deposition of metal films assisted by short wavelength laser irradiation [48]. The wave propagation along quartz surfaces in the middle infrared is also strongly modified by the reststrahlen properties of the quartz crystal, and dispersion effects due to this have been identified in the surface ripples written on quartz surfaces by TEA CO_2 laser pulses [55].

Several authors have noted in connection with these experiments that a surface polariton is tightly bound to the surface, and thus is more sensitive to the surface than the bulk properties of the medium. Turning the problem around, it has thus been suggested that the measured ripple spacings might be used to measure the exact dielectric constant of a surface layer [49], [71].

An interesting demonstration of much weaker surface dispersion effects has been provided in an elegant experiment by van Driel and co-workers [91]. Because of slight differences in the surface dielectric properties, ripples that are formed on germanium in the liquid phase where the entire surface is molten should have a spacing that differs by one percent or so from the spacing when the ripples are formed in the solid phase. By carefully observing the back-diffraction pattern of a visible probe beam while slowly increasing the peak power of laser pulses applied to a germanium surface, van Driel and co-workers were able to measure this small but significant difference in ripple spacing.

In another experiment performed on metals in the UV region, the ripple spacing was observed to change slowly during a multipulse illumination sequence as a thin oxide layer was formed and grew, thereby modifying the dielectric function at the surface and thus the dispersion of the surface wave, leading to a different spacing. The thickness of oxide layer could be estimated by such an experiment. Another strong dispersion effect observed in photodeposited grating growth on transparent substrates [125] is also discussed in the following subsection.

B. Ripple Growth Rates

The buildup from noise of stimulated Wood's anomalies, in common with other stimulated processes such as stimulated Brillouin or stimulated Raman scattering, can be expected to be more or less exponential, at least initially, with a very rapid growth rate, at least in those cases where there is a sharp single-pulse threshold and the laser energy fluence is above this threshold. We have noted, for example, that strong ripples can be formed on semiconductors using single laser pulses as short as 7 ps in duration. Observations of the exponential growth rate within the laser pulse in these cases are thus technically difficult.

Measurements of ripple growth rate in multiple-pulse cases where the ripples build up slowly over many repeated subthreshold pulses should be easier to implement since one can then monitor the ripple intensity versus number of repeated shots at low repetition rate and constant energy. Only a few studies of this sort have been reported [66]. It has been clearly demonstrated that the number of shots required to produce a given ripple strength increases rapidly as the pulse energy is reduced below the single-shot damage level.

The general characteristic of this type of ripple growth is that when a virgin surface is first illuminated with laser pulses at energies below the single-shot transformation energy, no readily visible ripple formation seems to occur during some finite number of initial pulses (although some sort of microdamage within the material must nonetheless be generated by these shots). Then, after a requisite number of "priming" pulses, visible ripple effects do occur and grow rapidly in strength with successive pulses up to some saturation level. The number of priming pulses required before visible ripples occur increases very rapidly as the pulse energy is lowered below, say, 50 percent of the single-shot threshold. This behavior is quite reminiscent of the statistical development of laser damage below the single-shot threshold, such as occurs in a large number of optical materials. Whether a definite lower threshold exists, below which no ripples will develop after any number of shots, seem to be, at present, uncertain.

One interesting technique for observing the resonant behavior of ripple growth rate versus ripple spacing in a photodeposition experiment has been demonstrated by Chen *et al.* [125], as illustrated in Fig. 10. In this experiment, two coherently related laser beams at 257 nm wavelength intersect at an experimentally variable angle

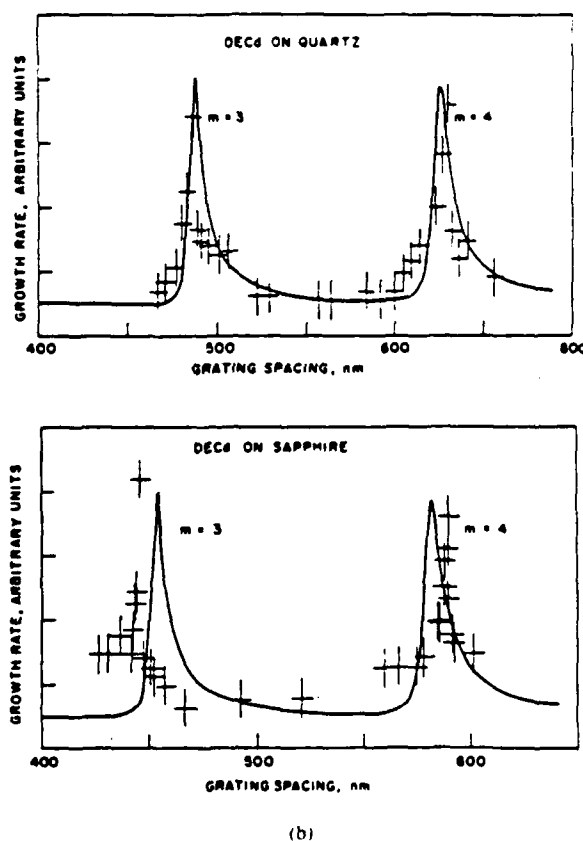
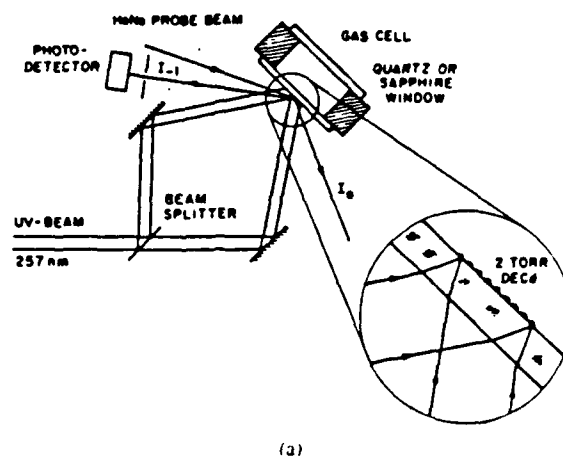


Fig. 10. (a) Experimental apparatus for photodeposition of Cd gratings on the inner surface of a cell wall, using two 257 nm UV beams with variable angle of interference. (b) Growth rates of the photodeposited cadmium gratings for two different substrates plotted versus grating spacing, which was varied by changing the angle of intersection between the two UV beams. From Chen *et al.* [125].

to form a sinusoidal intensity pattern with a variable period Λ on the inside surface of a quartz or sapphire cell filled with dimethyl- or diethyl-cadmium vapor. The ultraviolet radiation then dissociates metallorganic molecules that have been absorbed on the inside cell wall, forming a photodeposited cadmium grating with a transverse spatial variation matching the local light intensity.

Fig. 10 shows results for the measured growth rate

of such gratings versus the grating spacing as varied by changing the intersection angle between the two ultraviolet beams. The authors phrase their interpretation of these data in somewhat different terms. However, the observed peaks in growth rate for the photodeposited gratings correspond very closely to the Rayleigh conditions at which each of the two intersecting beams is tangentially diffracted along the cell surface in m th order. Strong peaks are seen for m ranging from 3 to 6 (only the first two of these peaks are shown in Fig. 10). Note that the Rayleigh grazing wave in this case travels primarily inside the transparent dielectric medium, and thus has a phase velocity characteristic of the dielectric substrate rather than the vapor above the substrate. Also, the intersecting beams in this case seem to have been polarized in TE rather than TM polarization.

C. Moving Ripple Growth

We pointed out in Section IV that with nonnormal incidence of the primary laser beam onto a sinusoidally corrugated surface, the sinusoidal interference fringes of the total Poynting vector into the surface may be shifted in spatial phase relative to the original surface ripples so that the peaks of the Poynting flux are neither directly in phase nor 180° out of phase with the surface corrugations. Hence, the optical energy deposition at the surface may have an arbitrary spatial phase shift with respect to the initial surface disturbance that produced the diffracted light and the Poynting vector modulation. This implies that with pulsed illumination, the maximum power deposition during each successive shot will not occur exactly at either the trough or the peak of the previously existing corrugations, but somewhere in between. This, in turn, may cause the ripples to move or slide forward or backward in the direction of the ripple k vector on successive shots.

A Russian group has now reported observation of this type of moving ripple behavior using pulsed illumination on germanium at $1.06\ \mu\text{m}$ [121]. By photographing the ripple patterns on a germanium surface containing several microscopic markers, they observed that the ripples did indeed slowly shift in position with repeated illumination in the direction parallel to the k vector of the previous corrugations. Both forward and backward motions of the ripples were seen at different angles of incidence of the primary laser beam.

D. Ripple Growth at Magic Angles

We noted in Section III that the tangential Rayleigh diffraction condition could be satisfied simultaneously in two different orders m_1 and m_2 at the same incident wavelength, corresponding to simultaneous tangential diffraction in both directions along the grating surface, provided the primary beam was incident at one of the magic angles given by Palmer [12]. One might then expect some significant change in ripple growth rate or in other characteristics of the ripples for a primary laser beam incident at or near one of these magic angles. Given the anti-cross-

ing behavior for linear Wood's anomalies near these angles, as reported by Stewart and Gallaway [17], however, it is not clear whether the ripple formation process might be enhanced or hindered. No experiments testing this point have yet been reported, although it should be an interesting experiment to perform.

E. CW Ripple Formation

One of the most interesting demonstrations of spontaneous ripple formation has been the observation of continuous or CW ripples on thin evaporated silicon films using continuous-wave $10.6\ \mu\text{m}$ radiation by Nemanich *et al.* [81] and later by Sipe and van Driel [143]. In these experiments, thin films of silicon evaporated onto quartz substrates were illuminated at normal incidence by the uniform beam from a low-power CW CO_2 laser, while simultaneously being observed from behind through the quartz substrate by a TV camera sensitive in the near infrared.

As the CO_2 laser intensity was slowly turned up through a certain threshold value, the silicon film was observed to suddenly separate into a regular grating of parallel molten and nonmolten stripes which could readily be observed by the IR camera because of their sharply different temperatures. The periodicity of these stripes corresponded to the $10.6\ \mu\text{m}$ wavelength of the CO_2 laser beam. Lowering the intensity caused the film to return to a uniform nonmolten state with no residual grating effects, and the system could be cycled repeatedly up and down through the grating formation threshold.

F. Crystal Structure Effects

Possible variations of the ripple formation process with crystal orientation on single-crystal samples have not been widely reported nor, perhaps, widely looked for. In experiments to be discussed in more detail below, however, we have noted that when an elliptical or circularly polarized, low-power pulsed $1.06\ \mu\text{m}$ laser beam strikes a $\langle 111 \rangle$ Ge surface at normal incidence, the multiple-pulse ripples grow initially with discrete grating k vectors of equal length that are symmetrically distributed in angle with six-fold symmetry in the surface plane [73]. The initial back-diffraction pattern thus appears to be six bright dots equally spaced around the central circular arc in Fig. 5(b). SEM photographs of the same surfaces reveal a definite, if somewhat irregular, pattern of hexagonal structures on the germanium surface. This symmetric growth pattern presumably results from a symmetry which is related to the crystal structure, either in the surface optical properties which control the ripple growth rate or in the initial surface noise disturbances from which the ripples first develop.

G. Higher Order Nonlinear Ripples

As we have noted earlier, interference between a primary laser beam and a scattered optical wave traveling at near-grazing incidence along a material surface will produce a periodically varying Poynting vector directed into

the surface. If the surface responds in some linear manner to the local Poynting flux directed into the surface, the result will be to write or emboss a grating pattern into the surface, with a grating vector k_g equal to the difference between the incident and scattered optical wavevector components in the surface plane. The period of the surface grating is given by $\Lambda = 2\pi/|k_g|$ where Λ is also the period of the optical intensity pattern on the surface.

In some cases, however, the material surface may respond *nonlinearly* to the sinusoidally varying light intensity. The result will then be to produce a grating having spatial components at the fundamental and various harmonics of the optical interference pattern so that a grating pattern with vector components $\pm nk_g$, where n takes on integer values, may be formed on the surface. In practical terms, this means that "harmonic" ripples with periods Λ/n may be formed on top of or along with the fundamental grating having period Λ . Ehrlich and Breuck [126] have used this "harmonic grating" process, for example, to grow 160 nm ripples on a surface having a preengraved 320 nm grating structure under varying conditions such that the incident 257.2 nm radiation was synchronous for stimulated grating growth either at 320 nm or, separately, at 160 nm.

Suppose further that a surface already has some periodic structure embossed upon it, with grating vector k_a , and that the surface is illuminated with an optical interference pattern having grating vector k_b —in other words, new ripples being written on top of old ripples. The effect of a nonlinear surface response will then be to produce additional intermodulation grating patterns with grating vectors given by $k_{ab} \equiv n_a k_a + n_b k_b$ where n_a and n_b are integers. Note that in this case, while the original k_a or k_b grating components may satisfy the Rayleigh condition for linear or stimulated Wood's anomalies at the wavelength in question, none of the higher order k_{ab} components may satisfy the Rayleigh condition for further growth of the stimulated ripples.

These ripples can, nonetheless, grow. Fig. 11(a) shows, for example, a photograph of the back-diffraction pattern from a (111) germanium surface after illumination by a large number of normally incident, circularly polarized, 1.06 μm pulses, illustrating the formation of second- and third-order nonlinear surface ripples [73]. In this experiment, the laser pulses striking the surface initially create first-order ripples which travel in preferred directions with six-fold symmetry on the crystal surface, leading to six equally spaced bright dots on the innermost circle in the figure. As the number of pulses incident on the surface increases, similar first-order ripples gradually build up in all other directions on the surface, eventually causing the entire circular arc in the back-diffraction pattern to be filled in.

At the same time, however, as new ripples are written on top of the existing ripples, higher order nonlinear ripples begin to appear, leading to the additional quasi-elliptical and hyperbolic ovals apparent in the figure. The diagram of Fig. 11(b) shows how the grating k vectors of

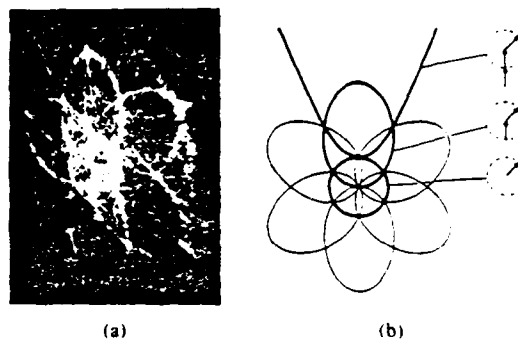


Fig. 11. (a) Back-diffraction pattern observed at 532 nm from a germanium (111) surface illuminated by a large number of low-intensity circularly polarized 1.06 μm pulses. (The original photograph is a color slide of the screen as seen from the sample side.) (b) Explanation of the observed diffraction patterns in terms of nonlinear sum-frequency gratings produced by the combination of one, two, or three grating k vectors.

these additional gratings are made up of linear sums of the first-order gratings running in various directions on the surface. In general, the nonlinear gratings are formed from the combination of either one or two components of the strong gratings along the preferred crystal directions, plus one additional grating vector whose orientation rotates through all directions in the surface plane.

H. Dynamic Ripple Behavior

As we mentioned above, the ripple growth rate during a single above-threshold laser pulse can be expected to be very fast, and thus difficult to observe directly. The principal time-resolved ripple formation experiments to date seem to be those of Keilmann [79], who observed the back-diffracted signals from a liquid mercury surface, and those of van Driel and co-workers [89]–[91], who observed the time-resolved back-diffraction signals at various angles from a germanium surface illuminated by 200 ns long 1.06 μm pulses of various intensities. van Driel's observations showed, however, not so much the dynamic growth rates for the ripples themselves (which were generally faster than the time resolution of the experiment), but rather interesting changes in the ripple amplitude versus time caused by changes (e.g., sudden melting or re-freezing) in the physical character of the surface itself. In particular, under conditions where the germanium surface was melted early in the laser pulse, subsequent strong periodic time variations in ripple amplitude were observed, which the authors attributed to stimulation of liquid capillary waves on the surface of the molten germanium, as proposed also by Keilmann [78], [79].

VI. APPLICATIONS

There is evidently a wide range of conditions under which spontaneous surface ripples can occur, and once such surface structures have been formed, they can have a strong influence on the optical properties of surfaces. As a result, the possible occurrence of stimulated (or linear) Wood's anomalies must be taken into account in a wide range of practical applications, including the following.

1) *Surface Physics Studies Using Low-Intensity Lasers:* As noted earlier, even when a surface is illuminated by comparatively low-intensity pulsed laser light, strong surface gratings may still build up slowly over many repeated shots. The development of such surface gratings may not be noticed if not carefully looked for, and their presence may substantially change the interaction of the laser beam with the surface, including the measured reflection and transmission properties of the surface. One should be on guard, therefore, in any physical or chemical experiments involving interaction of even low-intensity pulsed beams with surfaces, that stimulated Wood's anomalies are not present as a source of experimental error.

2) *Laser Annealing, Photoetching, and Laser Film Deposition:* The conditions required for laser annealing and similar kinds of photoprocessing of semiconductor surfaces, of course, involve exactly the kinds of higher laser intensities and optically induced surface phase changes which we have seen can readily lead to the formation of strong surface ripples. Practical techniques for laser-controlled surface etching and laser film deposition can similarly lead to formation of surface structures, even with comparatively low intensities and longer exposure times [112]. The generation of surface irregularities due to stimulated Wood's anomalies will usually be deleterious to the desired results in these applications, and the changes in surface optical properties that occur once these ripples have developed can lead to significant changes in the process characterization.

Techniques for avoiding the formation of surface ripples in these applications can include the use of circularly polarized light (at normal incidence) and possibly the use of two different, widely spaced optical wavelengths to accomplish the desired processing.

3) *Laser Materials Processing and Laser Damage Studies:* Other types of laser-assisted materials processing and machining, including laser cutting, welding heat treating, and annealing of many different materials, form another general area in which strong laser beams and major surface phase changes are present. The formation of spontaneous surface ripples is very often observed in these procedures, and again the change in surface properties caused by the ripples can produce major changes in the process variables. Studies of laser damage mechanisms can be similarly perturbed by unanticipated ripple formation effects.

4) *Single-Beam Fabrication of Optical Gratings:* Spontaneous surface ripples of the type described in this paper might potentially be employed as optical diffraction gratings, fabricated using a single laser beam rather than two interfering optical beams. The practical difficulties involved in controlling such a highly nonlinear process, achieving highly regular grating spacing over a large area, and controlling the final groove shape might seem to limit the practical utility of this technique compared to conventional techniques for fabricating high-quality holographic optical gratings.

Advantageous features, on the other hand, include the fact that the resulting ripple spacing can be substantially shorter than the primary laser wavelength, especially if one takes advantage of higher order scattering into surface waves seeded by initial preparation of a subharmonic grating on the surface. Uniform ripple spacing and phase can also be promoted over larger areas by stepping the laser beam by a fraction of the beam spot size in the direction of the grating k vector on successive shots so that diffraction from previously formed ripples seeds the growth of new ripples. Ehrlich and Brueck [126] have, for example, demonstrated the growth of a grating with 160 nm period using a primary laser wavelength of 257 nm on a surface which had been prepared with a seed grating at the subharmonic period of 320 nm. Other possibilities for single-beam fabrication of optical gratings have also been discussed by Aksenov [93].

5) *Enhanced Coupling to Photodetectors:* Finally, we might note that the enhanced surface coupling or reduced surface reflectivity produced by even the linear or conventional Wood's anomalies can be put to practical use, for example, in increasing the surface coupling of incident laser light and thereby increasing the detection sensitivity of solid-state photodetectors. Experimental results showing increased speed and sensitivity under Wood's anomaly conditions in photodetectors fabricated with deliberately periodic surface structures have been published both by Glass *et al.* [107] working with amorphous silicon photodetectors, by Brueck *et al.* [123], [124] who coupled into surface plasma waves in Au-coated InAs photodetectors, and by Berthod *et al.* [139] using holographic gratings on metal-insulator-semiconductor junctions in the visible.

ACKNOWLEDGMENT

A. E. Siegman is grateful for hospitality from the Max Planck Institute for Quantum Optics, Garching, West Germany, during 1984-1985.

REFERENCES

References on Wood's Anomalies

- [1] R. W. Wood, "On a remarkable case of uneven distribution of light in a diffraction grating spectrum," *Proc. Phys. Soc. (London)*, vol. 18, pp. 269-275, June 1902.
- [2] —, "On a remarkable case of uneven distribution of light in diffraction grating problems," *Phil. Mag.*, 6th ser., vol. 4, pp. 396-402, Sept. 1902.
- [3] —, "Diffraction gratings with controlled groove form and abnormal distribution of intensity," *Phil. Mag.*, 6th ser., vol. 23, pp. 310-317, Feb. 1912.
- [4] Lord Rayleigh, "On the dynamical theory of gratings," *Proc. Roy. Soc. (London)*, vol. A79, pp. 399-416, June 1907.
- [5] —, "Note on the remarkable case of diffraction spectra described by Prof. Wood," *Phil. Mag.*, 6th ser., vol. 14, pp. 60-65, July 1907.
- [6] L. R. Ingersoll, "Polarization of radiation by gratings," *Astrophys. J.*, vol. 51, pp. 129-139, Apr. 1920.
- [7] —, "Some peculiarities of polarization and energy distribution by speculum gratings," *Phys. Rev.*, vol. 17, pp. 493-501, Apr. 1921.
- [8] J. Strong, "Anomalous behavior of diffraction gratings coated by evaporation," *Phys. Rev.*, vol. 46, p. 326, Aug. 15, 1934.
- [9] —, "Polarization bands," *Phys. Rev.*, vol. 48, p. 480, Sept. 1, 1935.

- [10] —, "Effect of evaporated films on energy distribution in grating spectra," *Phys. Rev.*, vol. 49, pp. 291-296, Feb. 15, 1936.
- [11] R. W. Wood, "Anomalous diffraction gratings," *Phys. Rev.*, vol. 48, pp. 928-936, Dec. 15, 1935.
- [12] C. H. Palmer, Jr., "Parallel diffraction grating anomalies," *J. Opt. Soc. Amer.*, vol. 42, pp. 269-276, Apr. 1952.
- [13] V. Twersky, "On the scattering of waves by an infinite grating," *IRE Trans. Antennas Propagat.*, vol. AP-4, pp. 330-345, July 1956. (Review.)
- [14] —, "On multiple scattering of waves," *J. Res. Nat. Bur. Stand.*, vol. 64D, pp. 715-730, Nov.-Dec. 1960. (Review.)
- [15] R. F. Millar, "Scattering by a grating I," *Can. J. Phys.*, vol. 39, pp. 81-103, Jan. 1961. (Review.)
- [16] —, "Scattering by a grating II," *Can. J. Phys.*, vol. 39, pp. 104-118, Jan. 1961.
- [17] J. E. Stewart and W. S. Gallaway, "Diffraction anomalies in grating spectrophotometers," *Appl. Opt.*, vol. 1, pp. 421-429, July 1962.
- [18] A. Hessel and A. A. Oliner, "A new theory of Wood's anomalies on optical gratings," *Appl. Opt.*, vol. 4, pp. 1275-1297, Oct. 1965.
- [19] D. Y. Tseng, A. Hessel, and A. A. Oliner, "Scattering by a multimode corrugated structure with application to p type Wood anomalies," *Alta Freq.*, vol. 38N, Special Issue, p. 82, May 1969.
- [20] R. Petit, *Electromagnetic Theory of Gratings*. Heidelberg: Springer-Verlag, 1980.
- [21] A. A. Oliner, "Historical perspectives on microwave field theory," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 1022-1045, Sept. 1984.
- riodic surface structure in laser annealing," *Appl. Phys. Lett.*, vol. 35, pp. 782-784, Nov. 15, 1979.
- [36] G. A. Rozgonyi, H. J. Leamy, T. T. Sheng, and G. K. Celler, "The microstructure of laser annealed silicon," in *Laser-Solid Interactions and Laser Processing-1978*, S. D. Fernis, H. J. Leamy, and J. M. Poate, Eds., AIP Proc. 50, Amer. Inst. Phys., NY, 1979, pp. 457-462.
- [37] K. Affolter, W. Luthy, and M. Wittmer, "Interference effects on the surface of Nd:YAG-laser-reacted Pd-silicide," *Appl. Phys. Lett.*, vol. 36, pp. 559-561, Apr. 1, 1980. (Localized defect structures.)
- [38] M. F. Becker, R. M. Walser, J. G. Ambrose, and D. Y. Sheng, "Picosecond, 1.06 micron laser-induced amorphous phases in thin, single crystal silicon membranes," in *Picosecond Phenomena II*, R. M. Hochstrasser, W. Kaiser, and C. V. Shank, Eds., Springer-Verlag Ser. in Chem. Phys., 1980, pp. 290-293.
- [39] C. Hill and D. J. Godfrey, "Changes in surface topography after pulsed laser anneal of silicon," *J. Physique*, vol. 41, pp. 79-84, May 1980. (Localized defects.)

Ripples References, 1981

- [40] A. K. Jain, V. N. Kulkarni, D. K. Sood, and J. S. Uppal, "Periodic surface ripples in laser-treated aluminum and their use to determine absorbed power," *J. Appl. Phys.*, vol. 52, pp. 4882-4884, July 1981.
- [41] P. A. Temple and M. J. Soileau, "Polarization charge model for laser-induced ripple patterns in dielectric materials," *IEEE J. Quantum Electron.*, vol. QE-17, pp. 2067-2072, Oct. 1981.
- [42] J. A. Van Vechten, "Experimental tests for boson condensation and superconductivity in semiconductors during pulsed beam annealing," *Solid State Commun.*, vol. 39, pp. 1285-1291, Sept. 1981.
- [43] R. M. Walser, M. F. Becker, J. G. Ambrose, and D. Y. Sheng, "Heterogeneous nucleation of spatially coherent damage structures in crystalline silicon with picosecond 1.06 μm and 0.53 μm laser pulses," in *Laser and Electron-Beam Solid Interactions and Materials Processing*, J. F. Gibbons, L. D. Hess, and T. W. Sigmon, Eds., Amsterdam: Elsevier North-Holland, 1981, pp. 177-184.
- [44] J. S. Williams, W. L. Brown, G. K. Celler, H. J. Leamy, J. M. Poate, G. A. Rozgonyi, and T. T. Sheng, "Characterization of pulsed Nd:YAG laser-annealed, arsenic-ion-implanted silicon," *J. Appl. Phys.*, vol. 52, pp. 1038-1049, Feb. 1981.

Ripples References, 1982

- [45] V. P. Aksenov and B. G. Zhurkin, "Periodic structure formed during the action of high-power coherent radiation on a semiconductor surface," *Sov. Phys. Dokl.*, vol. 27, pp. 630-631, Aug. 1982.
- [46] S. I. Anisimov, S. M. Gol'berg, B. A. Malomed, and M. I. Tribel'skii, "Two-dimensional slightly supercritical structures in laser sublimation waves," *Sov. Phys. Dokl.*, vol. 27, pp. 130-132, Feb. 1982.
- [47] M. F. Becker, R. M. Walser, Y. K. Jhee, and D. Y. Sheng, "Picosecond laser damage mechanisms at semiconductor surfaces," *SPIE Proc.*, vol. 322, pp. 93-98, 1982.
- [48] S. R. J. Brueck and D. J. Ehrlich, "Simulated surface-plasma-wave scattering and growth of a periodic structure in laser photo-deposited metal films," *Phys. Rev. Lett.*, vol. 48, pp. 1678-1681, June 14, 1982.
- [49] D. J. Ehrlich, S. R. J. Brueck, and J. Y. Tsao, "Time-resolved measurements of stimulated surface polariton wave scattering and grating formation in pulsed-laser-annealed germanium," *Appl. Phys. Lett.*, vol. 41, pp. 630-632, Oct. 1, 1982. (Ge, time resolved)
- [50] P. M. Fauchet and A. E. Siegman, "Surface ripples on silicon and gallium arsenide under picosecond laser illumination," *Appl. Phys. Lett.*, vol. 40, pp. 824-826, May 1, 1982.
- [51] P. M. Fauchet, Z. Guosheng, and A. E. Siegman, "Periodic ripple structure on semiconductors under picosecond pulse illumination," in *Picosecond Phenomena III*, W. K. Kaiser, Ed., Heidelberg: Springer-Verlag, 1982, pp. 376-379.
- [52] A. M. Glass, P. E. Liao, D. H. Olson, and L. M. Humphrey, "Optical metal-oxide tunnel detectors with microstructured electrodes," *Opt. Lett.*, vol. 7, pp. 575-577, Dec. 1982. (Enhanced coupling to photodetectors.)
- [53] D. Haneman and R. J. Nemanich, "Surface topography of laser annealed silicon," *Solid State Commun.*, vol. 43, pp. 203-206, July 1982.

Ripples References, 1965-1970

- [22] M. Birnbaum, "Semiconductor surface damage produced by ruby lasers," *J. Appl. Phys.*, vol. 36, pp. 3688-3689, Nov. 1965.
- [23] M. Bertolotti, P. Mariotti, D. Sette, L. Stagni, and G. Vitali, "Electron microscope observation of laser damage of germanium," *Rad. Eff.*, vol. 1, pp. 161-172, July 1969.

Ripples References, 1971-1975

- [24] D. C. Emmony, R. P. Howson, and L. J. Willis, "Laser mirror damage in germanium at 10.6 μm ," *Appl. Phys. Lett.*, vol. 23, pp. 598-600, Dec. 1, 1973.
- [25] M. Siegrist, G. Kaech, and F. K. Kneubuhl, "Formation of a periodic wave structure on the dry surface of a solid by TEA-CO₂ laser pulses," *Appl. Phys.*, vol. 2, pp. 45-46, July 1973.
- [26] J. M. Elson and R. H. Ritchie, "Diffuse scattering and surface-plasmon generation by photons at a rough dielectric surface," *Phys. Status Solidi B*, vol. 62, pp. 461-468, Apr. 1, 1974.
- [27] C. T. Walters, "Surface scattering at LSD-wave initiation sites on nonmetallic materials," *Appl. Phys. Lett.*, vol. 25, pp. 696-698, Dec. 15, 1974.
- [28] D. C. Emmony, N. J. Phillips, J. H. Toyer, and L. J. Willis, "The topography of laser-irradiated germanium," *J. Phys. D (Appl. Phys.)*, vol. 8, pp. 1472-1479, Sept. 11, 1975.
- [29] L. J. Willis and D. C. Emmony, "Laser damage in germanium," *Opt. Laser Technol.*, vol. 7, pp. 222-228, Oct. 1975.

Ripples References, 1976-1980

- [30] J. C. Koo and R. E. Slusher, "Diffraction from laser-induced deformation on reflective surfaces," *Appl. Phys. Lett.*, vol. 28, pp. 614-616, May 15, 1976.
- [31] P. A. Temple and M. J. Soileau, "Resonant defect enhancement of the laser electric field," in *Laser Induced Damage in Optical Materials*, A. J. Glass and A. H. Guenther, Eds., NBS Special Publ. 462, 1976, pp. 371-378.
- [32] N. R-Asenor, "CO₂ laser-produced ripple patterns on Ni, Pt, surfaces," *Appl. Phys. Lett.*, vol. 31, pp. 148-150, Aug. 1, 1977.
- [33] H. J. Leamy, G. A. Rozgonyi, T. T. Sheng, and G. K. Celler, "Periodic regrowth phenomena produced by laser annealing of ion-implanted silicon," *Appl. Phys. Lett.*, vol. 32, pp. 535-537, May 1, 1978.
- [34] G. N. Maracas, G. L. Harris, C. A. Lee, and R. A. McFarlane, "On the origin of periodic surface structure of laser-annealed semiconductors," *Appl. Phys. Lett.*, vol. 33, pp. 453-455, Sept. 1, 1978. (Ripples with unusual spacing.)
- [35] M. Oron and G. Sorensen, "New experimental evidence of the pe-

- [54] M. Hasselbeck and H. S. Kwok, "CO₂ laser-induced melting of indium antimonide," *Appl. Phys. Lett.*, vol. 41, pp. 1138-1140, Dec. 15, 1982.
- [55] F. Keilmann and Y. H. Bai, "Periodic surface structures frozen into CO₂ laser-melted quartz," *Appl. Phys. A*, vol. 29, pp. 9-18, Sept. 1982.
- [56] R. M. Osgood, Jr., and D. J. Ehrlich, "Optically induced microstructures in laser-photodeposited metal films," *Opt. Lett.*, vol. 7, pp. 385-387, Aug. 1982. (Photodeposition.)
- [57] A. M. Prokhorov, V. A. Sychugov, A. V. Tishchenko, and A. A. Khakimov, "Excitation of surface electromagnetic waves at a solid surface by an intense laser beam," *Sov. Tech. Phys. Lett.*, vol. 8, pp. 415-416, Aug. 1982.
- [58] M. J. Soileau and E. W. Van Stryland, "Ripple structure and enhanced absorption associated with ordered surface defects," in *Proc. 14th Annu. Symp. Opt. Materials for High Power Lasers*, Boulder, CO, 1982.
- [59] M. I. Tribel'skii and S. M. Gol'berg, "Resonant excitation of electromagnetic surface waves in laser evaporation of condensed media," *Sov. Tech. Phys. Lett.*, vol. 8, pp. 526-528, Oct. 1982.
- [60] H. M. van Driel, J. E. Sipe, and J. F. Young, "Laser-induced periodic surface structure on solids: A universal phenomenon," *Phys. Rev. Lett.*, vol. 49, pp. 1955-1958, Dec. 27, 1982.
- [61] J. F. Young, J. E. Sipe, M. I. Gallant, J. S. Preston, and H. M. van Driel, "Periodic surface damage in germanium: The spectrum in k space and its theoretical interpretation," in *Laser and Electron-Beam Interactions with Solids*, B. R. Appleton and G. K. Celler, Eds., Amsterdam: Elsevier, 1982, pp. 233-237.
- [62] J. F. Young, J. E. Sipe, J. S. Preston, and H. M. van Driel, "Laser-induced periodic surface damage and radiation remnants," *Appl. Phys. Lett.*, vol. 41, pp. 261-264, Aug. 1, 1982.
- [63] G. N. Zhizhin, M. A. Moskalova, E. V. Shomina, and V. A. Yakovlev, "Surface electromagnetic wave propagation on metal surfaces," in *Surface Polaritons*, V. M. Agranovich and D. L. Mills, Eds., New York: North-Holland, 1982.
- [64] G. N. Zhizhin, A. A. Sigarev, V. A. Sychugov, and A. A. Khakimov, "Resonant conversions of surface electromagnetic waves at a rippled metal surface," *Sov. Tech. Phys. Lett.*, vol. 8, pp. 214-216, Apr. 1982.
- [65] Z. Guosheng, P. M. Fauchet, and A. E. Siegman, "Growth of spontaneous periodic surface structures on solids during laser illumination," *Phys. Rev. B*, vol. 26, pp. 5366-5381, Nov. 15, 1982.
- [66] P. M. Fauchet, "Gradual surface transitions on semiconductors induced by multiple picosecond laser pulses," *Phys. Lett.*, vol. 93A, p. 155, 1983.
- [67] P. V. Bazukutsa, A. M. Prokhorov, V. A. Sychugov, and A. V. Tishchenko, "Surface state of germanium and its reaction to intense laser bombardment," *Sov. Tech. Phys. Lett.*, vol. 9, pp. 234-235, May 1983.
- [68] V. V. Bazhenov, A. M. Bonch-Bruевич, M. N. Libenson, V. S. Makin, S. D. Pudkov, and V. V. Trubaev, "Periodic structures formed by intense electromagnetic radiation on the surface of a semiconductor," *Sov. Tech. Phys. Lett.*, vol. 9, pp. 402-404, Aug. 1983.
- [69] —, "Dispersion of surface electromagnetic waves on semiconductors with a periodic surface relief caused by intense irradiation," *Sov. Tech. Phys. Lett.*, vol. 9, pp. 544-545, Oct. 1983.
- [70] D. O. Boerma, H. Hasper, and K. G. Prasad, "Evaporation and ripple formation during pulsed laser irradiation of GaAs," *Phys. Lett.*, vol. 93A, pp. 253-256, Jan. 17, 1983.
- [71] D. J. Ehrlich, S. R. J. Brueck, and J. Y. Tsao, "Surface electromagnetic waves in laser material interactions," in *Mater. Res. Soc. Symp. Proc.*, vol. 13, 1983, pp. 191-196.
- [72] V. I. Emel'yanov, E. M. Zemskov, and V. N. Seminogov, "Theory of formation of surface gratings under the action of laser radiation on surfaces of metals, semiconductors, and insulators," *Sov. J. Quantum Electron.*, vol. 13, pp. 1556-1561, Dec. 1983.
- [73] P. M. Fauchet and A. E. Siegman, "Observation of higher-order laser-induced surface ripples on (111) germanium," *Appl. Phys. A*, vol. 32, pp. 135-140, Nov. 1983.
- [74] P. M. Fauchet, G. S. Zhou, and A. E. Siegman, "Picosecond laser-induced surface transformation in solids," in *Laser-Solid Interactions and Transient Thermal Processing of Materials*, J. Narayan, W. L. Brown, and R. A. Lemons, Eds., New York: North-Holland, 1983, pp. 205-210.
- [75] J. F. Figueira and S. J. Thomas, "Generation of surface microstructures in metals and semiconductors by short pulse CO₂ lasers," in *Surface Studies with Lasers*, F. R. Aussenegg, A. Leitner, and M. E. Lippitsch, Eds., Berlin: Springer-Verlag, 1983, pp. 212-215.
- [76] G. M. Gandel'man and P. S. Kondratenko, "Complete suppression of metallic reflection upon the resonant excitation of surface plasma waves," *JETP Lett.*, vol. 38, pp. 291-293, Sept. 10, 1983.
- [77] G. G. Gromov and V. B. Ufimtsev, "Laser-induced two-dimensional lattice on an InSb surface," *Sov. Tech. Phys. Lett.*, vol. 9, pp. 249-250, May 1983.
- [78] F. Keilmann, "Instability of liquid metal surfaces under intense infrared irradiation," in *Surface Studies with Lasers*, F. R. Aussenegg, A. Leitner, and M. E. Lippitsch, Eds., Berlin: Springer-Verlag, 1983, pp. 207-211; see also *J. Physique Colloque C5*, vol. 44, pp. C5-77-C5-78, Oct. 1983.
- [79] F. Keilmann, "Laser-driven corrugation instability of liquid metal surfaces," *Phys. Rev. Lett.*, vol. 51, pp. 2097-2100, Dec. 5, 1983.
- [80] R. J. Nemanich, D. K. Biegelsen, and W. G. Hawkins, "Aligned, coexisting liquid and solid regions in laser-annealed Si," *Phys. Rev.*, vol. B27, pp. 7817-7819, June 15, 1983.
- [81] —, "Aligned, coexisting liquid and solid regions in pulsed and cw laser annealing of Si," in *Mater. Res. Soc. Symp. Proc.*, vol. 13, 1983, pp. 211-216.
- [82] A. M. Prokhorov, A. S. Svakhin, V. A. Sychugov, A. V. Tishchenko, and A. A. Khakimov, "Excitation and resonant transformation of a surface electromagnetic wave during irradiation of a solid by high-power laser radiation," *Sov. J. Quantum Electron.*, vol. 13, pp. 568-571, May 1983.
- [83] A. M. Prokhorov, V. A. Sychugov, A. V. Tishchenko, and A. A. Khakimov, "Resonant conversions of surface electromagnetic waves on a germanium surface bombarded by an intense laser beam," *Sov. Tech. Phys. Lett.*, vol. 9, pp. 28-29, Jan. 1983.
- [84] A. A. Samokhin, V. A. Sychugov, and A. V. Tishchenko, "Mechanisms of formation of periodic structures under the action of radiation on absorbing condensed media," *Sov. J. Quantum Electron.*, vol. 13, pp. 1433-1435, Oct. 1983.
- [85] A. A. Samokhin, V. A. Sychugov, A. V. Tishchenko, and A. A. Khakimov, "Characteristics of formation of a periodic structure due to absorption of monochromatic radiation by the surface of a solid," *Sov. J. Quantum Electron.*, vol. 13, pp. 659-660, May 1983.
- [86] J. E. Sipe, J. F. Young, J. S. Preston, and H. M. van Driel, "Laser-induced periodic surface structure. I. Theory," *Phys. Rev.*, vol. B27, pp. 1141-1154, Jan. 15, 1983.
- [87] N. Tsukada, S. Sugata, and Y. Mita, "New experimental evidence of surface ripples on gallium arsenide in laser annealing," *Appl. Phys. Lett.*, vol. 42, pp. 424-426, Mar. 1, 1983.
- [88] N. Tsukada, S. Sugata, H. Saitoh, and Y. Mita, "Surface ripples in laser-photochemical wet etching of gallium arsenide," *Appl. Phys. Lett.*, vol. 43, pp. 189-191, July 15, 1983.
- [89] J. F. Young, J. S. Preston, J. E. Sipe, and H. M. van Driel, "Time-resolved evolution of laser-induced periodic surface structure on germanium," *Phys. Rev.*, vol. B27, pp. 1424-1429, Jan. 15, 1983.
- [90] J. F. Young, J. S. Preston, H. M. van Driel, and J. E. Sipe, "Laser-induced periodic surface structure. II. Experiments on Ge, Si, Al, and brass," *Phys. Rev.*, vol. B27, pp. 1155-1172, Jan. 15, 1983.
- [91] J. F. Young, J. E. Sipe, and H. M. van Driel, "Regimes of laser-induced periodic surface structure on germanium: Radiation remnants and surface plasmons," *Opt. Lett.*, vol. 8, pp. 431-433, Aug. 1983.
- [92] H. M. van Driel, J. F. Young, and J. E. Sipe, "Laser-induced periodic surface structure," in *Mater. Res. Soc. Symp. Proc.*, vol. 13, 1983, pp. 197-203.

Ripples References, 1983

Ripples References, 1984

- [93] V. P. Aksenov, "Single beam method of manufacturing holographic optical gratings," *Symp. Optika '84, Proc. SPIE*, vol. 473, pp. 294-297, Apr. 1984. (Grating fabrication.)
- [94] I. A. Avrutskii, P. V. Bazukutsa, A. M. Prokhorov, V. A. Sychugov, and A. V. Tishchenko, "Self-deepening of the dielectric constant modulation at a germanium surface in an intense light wave," *Sov. Tech. Phys. Lett.*, vol. 10, pp. 461-463, Sept. 1984.
- [95] I. A. Avrutskii, P. V. Bazukutsa, V. A. Sychugov, and A. V. Tishchenko, "Periodic microrelief on a germanium surface and the thermophysical mechanism for its formation," *Sov. Tech. Phys. Lett.*, vol. 10, pp. 563-565, Nov. 1984.
- [96] P. V. Bazukutsa, V. L. Maslennikov, A. M. Prokhorov, V. A. Sy-

- chugov, and A. V. Tishchenko, "Diffraction mechanism of periodic structure formation under the action of radiation on an absorbing condensed medium," *Sov. J. Quantum Electron.*, vol. 14, pp. 976-980, July 1984.
- [97] P. V. Bazakutsa, V. A. Sychugov, and A. M. Prokhorov, "Diffraction of light by surface irregularities and its role in the formation of a periodic surface microrelief," *Sov. J. Quantum Electron.*, vol. 14, pp. 1420-1421, Oct. 1984.
- [98] V. V. Bazhenov, A. M. Bonch-Bruевич, M. N. Libenson, and V. S. Makin, "Interference of surface electromagnetic waves in connection with periodic structures formed during intense illumination of a semiconductor surface," *Sov. Tech. Phys. Lett.*, vol. 10, pp. 642-645, Dec. 1984.
- [99] A. M. Bonch-Bruевич, M. N. Libenson, and V. S. Makin, "Growth dynamics of periodic surface structures during intense illumination of a condensed medium," *Sov. Tech. Phys. Lett.*, vol. 10, pp. 1-3, Jan. 1984.
- [100] I. W. Boyd, S. C. Moss, T. F. Boggess, and A. L. Smirl, "Initial observation of the crystal-amorphous transition and the formation of ripple patterns on silicon induced by 7 ps pulses at 1.05 μm ," in *Mater. Res. Soc. Symp. Proc.*, vol. 23, 1984, pp. 203-208.
- [101] —, "Various phase transitions and changes in surface morphology of crystalline silicon induced by 4-260 ps pulses of 1 μm radiation," *Appl. Phys. Lett.*, vol. 45, pp. 80-83, July 1, 1984.
- [102] M. Combescot, J. Bok, and C. Benoit la Guillaume, "Instability at the melting threshold of laser irradiation of silicon as the underlying origin of the ripples formation," in *Mater. Res. Soc. Symp. Proc.*, vol. 23, 1984, pp. 199-202.
- [103] V. I. Emel'yanov and V. N. Seminogov, "Laser excitation of coupled surface electromagnetic and acoustic waves and of static surface structures in solids," *Sov. Phys. JETP*, vol. 59, pp. 598-604, Mar. 1984.
- [104] —, "Excitation of coupled capillary waves, diffracted electromagnetic waves, and surface structures as a result of interaction of high-power electromagnetic radiation with liquid metals, semiconductors, and insulators," *Sov. J. Quantum Electron.*, vol. 14, pp. 591-592, May 1984.
- [105] V. I. Emel'yanov, E. M. Zemskov, and V. N. Seminogov, "Theory of the formation of normal and anomalous gratings on the surfaces of absorbing condensed media exposed to laser radiation," *Sov. J. Quantum Electron.*, vol. 14, pp. 1515-1521, Nov. 1984.
- [106] P. M. Fauchet and A. E. Siegman, "Surface damage mechanisms in nontransparent media," in *Proc. 16th Annu. Symp. Opt. Mater. for High Power Lasers*, Boulder, CO, 1984.
- [107] A. M. Glass, P. F. Liao, A. M. Johnson, L. M. Humphrey, R. Lemons, D. H. Olson, and M. B. Stern, "Periodically structured amorphous silicon detectors with improved picosecond responsivity," *Appl. Phys. Lett.*, vol. 44, pp. 77-79, Jan. 1, 1984. (Enhanced coupling to photodetectors.)
- [108] G. A. Golubenko, A. A. Samokhin, and V. A. Sychugov, "Formation of a periodic structure on the surface of a liquid metal by laser radiation," *Sov. J. Quantum Electron.*, vol. 14, pp. 1239-1240, Sept. 1984.
- [109] A. V. Gorbunov and N. V. Klassen, "Periodic damage of transparent insulators by CO₂-laser pulses," *Phys. Chem. Mech. Surfaces*, vol. 2, pp. 1117-1124, 1984.
- [110] Y. Kanemitsu, H. Kuroda, and S. Shionoya, "Formation of periodic ripple structures in picosecond pulsed laser annealing of ion-implanted silicon," *Japan. J. Appl. Phys.*, vol. 23, pp. 1060-1064, Aug. 1984.
- [111] N. Mansour, G. Reali, P. Aiello, and M. J. Soileau, "Laser generated ripple patterns on dielectrics and intermediate band gap semiconductors," in *Proc. 16th Annu. Symp. Opt. Mater. for High Power Lasers*, Boulder, CO, 1984.
- [112] A. M. Prokhorov, V. A. Sychugov, and T. V. Tulaikova, "Influence of a liquid waveguide on formation of gratings by photoetching of semiconductors," *Sov. Phys. Semicond.*, vol. 18, pp. 1325-1328, Dec. 1984. (Ripple formation during photoetching.)
- [113] M. J. Soileau, "Ripple structures associated with ordered surface defects in dielectrics," *IEEE J. Quantum Electron.*, vol. QE-20, pp. 464-467, May 1984.
- [114] V. A. Sychugov and T. V. Tulaikova, "Waveguide effect during photoetching of semiconductors," *Sov. J. Quantum Electron.*, vol. 14, pp. 301-302, Mar. 1984.
- [115] I. Ursu, I. N. Mihailescu, A. Popa, A. M. Prokhorov, V. I. Konov, V. P. Ageev, and V. N. Tokarev, "CO₂-laser radiation absorption by metal gratings," *Appl. Phys. Lett.*, vol. 45, pp. 365-367, Aug. 15, 1984.
- [116] V. P. Veiko, I. A. Dorofeev, Ya. A. Imas, T. I. Kalugina, M. N. Libenson, and G. D. Shandybina, "Formation of periodic structures on a silicon surface by a millisecond Nd-laser pulse," *Sov. Tech. Phys. Lett.*, vol. 10, pp. 6-8, Jan. 1984.
- [117] V. P. Veiko and E. A. Tuchkova, "Irregularity of the destruction front in a metal surface exposed to laser radiation," *Phys. Chem. Mech. Surfaces*, vol. 2, pp. 1260-1276, 1984.
- [118] G. Vitali, M. Mannelli, U. Zammit, and F. Scudieri, "Temperature gradient effects in low-power laser annealing of ion-implanted α -Ge," *Appl. Phys.*, vol. A35, pp. 233-239, Dec. 1984.
- [119] R. Wilson, F. A. Houle, and C. R. Jones, "Chemical composition variations across surface ripples in photodeposited films," in *Laser Chemical Processing of Semiconductor Devices*, Mater. Res. Soc. Symp. B, Boston, MA, 1984, pp. 90-91.
- [120] J. F. Young, J. E. Sipe, and H. M. van Driel, "Laser-induced periodic surface structure. III. Fluence regimes, the role of feedback, and details of the induced topography in germanium," *Phys. Rev. B*, vol. 30, pp. 2001-2015, Aug. 15, 1984.

Ripples References, 1985

- [121] I. A. Avrutskii, P. V. Bazakutsa, A. M. Prokhorov, and V. A. Sychugov, "Motion of a periodic surface microrelief under the action of high-power laser radiation," *Sov. J. Quantum Electron.*, vol. 15, pp. 429-430, Mar. 1985.
- [122] I. W. Boyd, S. C. Moss, T. F. Boggess, and A. L. Smirl, "Temporally resolved imaging of silicon surfaces melted with intense picosecond 1- μm laser pulses," *Appl. Phys. Lett.*, vol. 46, pp. 366-368, Feb. 15, 1985.
- [123] S. R. J. Brueck, V. Diadiuk, T. Jones, and W. Lenth, "Enhanced quantum efficiency internal photoemission detectors by grating coupling to surface plasma waves," *Appl. Phys. Lett.*, vol. 46, pp. 915-917, May 15, 1985. (Enhanced coupling to photodetectors.)
- [124] —, "High-speed internal photoemission detectors enhanced by grating coupling to surface plasma waves," in *Picosecond Electronics and Optoelectronics*, D. H. Bloom and G. Mourou, Eds. Berlin: Springer-Verlag, 1985. (Enhanced coupling to photodetectors.)
- [125] C. J. Chen, H. H. Gilgen, and R. M. Osgood, "Resonant optical growth of submicrometer metal gratings," *Opt. Lett.*, vol. 10, pp. 173-175, Apr. 1985. ("Transverse ripples.")
- [126] D. J. Ehrlich and S. R. J. Brueck, "Laser photochemical fabrication of phase-controlled 160-nm period gratings by stimulated second-order surface plasma wave scattering," *Appl. Phys. Lett.*, vol. 47, pp. 216-218, Aug. 1, 1985. (Grating fabrication: higher order ripples.)
- [127] P. M. Fauchet and A. E. Siegman, "Laser-induced surface ripples: What is understood and what is not," in *Mater. Res. Soc. Symp. Proc.*, vol. 35, 1985, pp. 199-204.
- [128] G. A. Golubenko, A. A. Samokhin, V. A. Sychugov, and A. V. Tishchenko, "Total reflection of light from a corrugated surface of a dielectric waveguide," *Sov. J. Quantum Electron.*, vol. 15, pp. 886-887, 1985.
- [129] K. W. Goossen and S. A. Lyon, "Grating enhanced quantum well detector," *Appl. Phys. Lett.*, vol. 47, pp. 1257-1259, Dec. 15, 1985.
- [130] G. Gorodetsky, J. Kanicki, T. Kazyaka, and R. L. Melcher, "Far UV pulsed laser melting of silicon," *Appl. Phys. Lett.*, vol. 46, pp. 547-549, Mar. 15, 1985.
- [131] R. B. James, J. Narayan, R. F. Wood, D. K. Ottesen, and K. F. Siegfriedt, "Optical studies during pulsed CO₂ laser irradiation of ion-implanted silicon," *J. Appl. Phys.*, vol. 57, pp. 4727-4731, May 15, 1985.
- [132] B. I. Makshantsev and N. F. Pilipetskii, "Formation of periodic structures on the surfaces of solids under the action of laser radiation," *Sov. J. Quantum Electron.*, vol. 15, pp. 563-566, Apr. 1985.
- [133] —, "On the formation of periodic structures on solid surfaces," *Appl. Phys.*, vol. A36, pp. 205-207, Apr. 1985.
- [134] F. Shaapur and S. D. Allen, "Periodic surface structures in laser chemical vapor deposited SiC," in *Mater. Res. Soc. Symp. D*, Boston, MA, Dec. 1985, paper D4-4.
- [135] I. Ursu *et al.*, "Energy plasma-target coupling in the case a periodical structure is formed within the irradiation spot as an effect of powerful laser radiation," *Opt. Commun.*, vol. 55, pp. 409-412, 1985.
- [136] I. Ursu, I. N. Mihailescu, L. C. Nistor, V. S. Teodorescu, A. M.

- Prokhorov, V. I. Konov, and V. N. Tokarev, "Periodic structures on the surface of fused silica under multipulse 10.6- μ m laser irradiation," *Appl. Opt.*, vol. 24, pp. 3736-3739, Nov. 15, 1985. (Ripple growth via surface vaporization.)
- [137] I. Ursu, I. N. Mihailescu, A. M. Prokhorov, V. I. Konov, and V. N. Tokarev, "On the role of the periodical structures induced by powerful laser irradiation of metal surfaces in the energy coupling process," *Physica*, vol. 132, pp. 395-402, Sept. 1985.
- [138] R. J. Wilson and F. A. Houle, "Composition, structure, and electric field variations in photodeposition," *Phys. Rev. Lett.*, vol. 55, pp. 2184-2187, Nov. 11, 1985.

Ripples References, 1986

- [139] K. Berthold, W. Beinstingl, R. Berger, and E. Gornik, "Surface plasmon enhanced quantum efficiency of metal-insulator-semiconductor junctions in the visible," *Appl. Phys. Lett.*, vol. 48, pp. 526-528, Feb. 24, 1986.
- [140] S. R. J. Brueck and D. J. Ehrlich, "Stimulated surface plasma wave scattering," in *Proc. 1st Int. Laser Sci. Conf.*, Dallas, TX, 1985.
- [141] P. M. Fauchet, "Properties of gratings written with a single laser beam," in *Proc. 1st Int. Laser Sci. Conf.*, Dallas, TX, 1985.
- [142] —, "Raman microprobe analysis of laser-induced microstructures," in *Beam-Solid Interactions and Phase Transformations*, H. Kurz, G. L. Olson, and J. M. Poate, Eds., Mater. Res. Soc., to be published.
- [143] J. H. Sipe and H. M. van Driel, "Laser induced surface gratings," in *Proc. 1st Int. Laser Sci. Conf.*, Dallas, TX, 1985.



Anthony E. Siegman (S'54-M'57-F'66) was born in Detroit, MI, on November 23, 1931. He received the A.B. degree summa cum laude from Harvard College, Cambridge, MA, in 1952, and the M.S. degree in applied physics from the University of California, Los Angeles, in 1954 under the Hughes Aircraft Company Cooperative Plan. In 1957 he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

Since then he has been on the faculty at Stanford, as a Professor of Electrical Engineering and (by courtesy) Applied Physics, where he directs a research program in lasers and their applications. During 1965 he served as Visiting Professor of Applied Physics at Harvard University, and in 1969-1970 he worked at the IBM Research Laboratory, Zurich, under a Guggenheim Fellowship. From 1978 to 1983

he was Director of the Edward L. Ginzton Laboratory at Stanford. He has made many contributions to the fields of microwave electronics and traveling-wave tubes, and to laser devices and their applications, and has written *Microwave Solid-State Masers* (McGraw-Hill, 1964), *An Introduction to Lasers and Masers* (McGraw-Hill, 1970) and *Lasers* (University Science Books, in preparation), as well as co-editing *Laser Devices and Applications* (IEEE Press, 1973). He holds four patents on microwave and optical devices and lasers. He was Program Chairman for the 1966 International Quantum Electronics Conference, and Conference Chairman for the 1968 IQEC. In November 1981 he presented the Annual Benjamin W. Lee Memorial Lectures on Physics in Seoul, Korea. In 1983 he served as Co-Director of the Far Eastern Laser School held in Korea, and in 1984 as Co-Director of the Winter School on Lasers and Laser Optics in Taiwan, R.O.C. He has also served as consultant to numerous industrial and government laboratories, and recently completed a term on the Air Force Scientific Advisory Board. He has been awarded a Senior U. S. Scientist Award by the Humboldt Foundation for 1984, and was on sabbatical leave at the Max-Planck-Institute in Munich, Germany, for the academic year 1984-1985. He has recently been elected a Fellow of the American Academy of Arts and Sciences.

Professor Siegman is a Fellow of the Optical Society of America, the American Physical Society, and the American Association for the Advancement of Science, and in 1973 was elected to the National Academy of Engineering. He is also a member of the American Association of University Professors, Phi Beta Kappa and Sigma Xi. In 1972, together with D. J. Kuizenga, he received the W. R. G. Baker Award of the IEEE for the best paper published in any IEEE TRANSACTIONS during 1971, and 1977 he received the J. J. Ebers Award of the Institute of Electrical and Electronics Engineers. In 1980 he received the R. W. Wood Prize of the Optical Society of America for the invention of the unstable optical resonator.

Philippe M. Fauchet received the Engineer's degree from the Faculté Polytechnique de Mons, Mons, Belgium, the M.S. degree from Brown University, Providence, RI, and the Ph.D. degree in applied physics from Stanford University, Stanford, CA.

After a year as an IBM Postdoctoral Fellow in the Edward L. Ginzton Laboratory at Stanford, in 1984 he joined the faculty of Princeton University, Princeton, NJ, where he is an Assistant Professor of Electrical Engineering. His research interests are in the fields of ultrafast optoelectronics and spectroscopy of semiconductors, laser-solid interactions including annealing and damage, and characterization of novel semiconductor microstructures. He has many publications in these fields.

Dr. Fauchet's work has been recognized by several prizes and awards, including those from the Materials Research Society and IBM.

A Reprint from the

PROCEEDINGS

Of SPIE - The International Society for Optical Engineering



Volume 793

**Ultrafast Lasers Probe Phenomena in Bulk
and Microstructure Semiconductors**

25-26 March 1987
Bay Point, Florida

Recombination dynamics in microstructures

J. E. Fouquet

Instruments and Photonics Laboratory, Hewlett Packard Laboratories
3500 Deer Creek Road, Palo Alto, California 94304

Invited Paper

Recombination Dynamics in Microstructures

J.E. Fouquet

Instruments and Photonics Laboratory, Hewlett Packard Laboratories
3500 Deer Creek Road, Palo Alto, California 94304

Abstract

Recombination dynamics in III-V quantum well structures have been investigated using a wide variety of optical techniques. Results of these studies have a number of implications for the design of light-emitting devices.

Introduction

Early successes in GaAs/Al_xGa_{1-x}As quantum well (QW) lasers sparked widespread interest in recombination dynamics in QW structures, in particular whether dynamics in III-V QW structures are more favorable for laser and light-emitting diode (LED) operation than those in related bulk semiconductors.

In order to design an optimal emitter in these structures, dominant recombination mechanisms must be identified and quantified. Studies of recombination in bulk III-V semiconductors¹ and early work in related QW structures^{2,3} suggest four possible radiative recombination mechanisms: bimolecular recombination, excitonic recombination, stimulated emission, and phonon-assisted stimulated emission. Possible nonradiative recombination mechanisms include trapping and, in long (>1 micron) wavelength emitters, Auger recombination. Some areas which have been particularly important to QW emitters are: (1) Is LED emission governed by bimolecular radiative recombination or by excitonic recombination? (2) If bimolecular recombination dominates, what is its rate? (3) Does phonon-assisted stimulated emission exist? (4) How serious is trapping in these structures? (5) How large is Auger recombination in long-wavelength emitters?

In order to answer these questions a number of experimental techniques were employed. Conventional photoluminescence (PL) at room temperature and liquid helium temperatures was used to investigate GaAs/Al_xGa_{1-x}As and Ga_{0.47}In_{0.53}As/InP QW structures. Time-correlated single photon counting was used to study time-resolved PL of GaAs/AlGaAs QW structures. Both CW and time-resolved PL were studied as a function of excitation density. Finally, electroluminescence from GaInAs/InP QW LEDs was observed as a function of current.

Bimolecular versus Excitonic Radiative Recombination

In excitonic recombination, an electron and hole are initially bound together by Coulomb interaction as an exciton. When they recombine radiatively, the photon emitted has an energy less than the $n=1$ electron to $n=1$ heavy hole QW transition energy by the Coulomb binding energy (typically ≤ 10 meV in GaAs/AlGaAs QWs). Optical emission depends on the population of excitons, which is proportional to the number of excitation photons. The emission rate should be relatively rapid ($\sim 10^9$ sec⁻¹)⁴ due to the good overlap of the electron and hole wavefunctions. In bimolecular radiative recombination, optical emission depends on the product of the free electron density and the free hole density; this product is proportional to the square of the number of excitation photons, since each photon generates an electron and a hole. Thus the bimolecular emission rate varies with carrier density but is generally slower than the excitonic rate because of poorer overlap of the electron and hole wave functions except at very high carrier densities. If excitonic recombination were dominant in emitters rather than bimolecular radiative recombination, radiative recombination would compete more effectively with nonradiative recombination and thus emitters would be brighter and faster. Unfortunately, this is not the case. At the carrier densities required for LEDs, nearly all recombination takes place between free carriers, not excitons.

Although excitons are abundant in both bulk GaAs and GaAs/AlGaAs QWs at very low temperatures, the binding energy of these excitons is less than kT at room temperature, leading to rapid ionization of excitons into free carriers at room temperature. This ionization rate has been measured to be ~ 1 psec.⁵ Excitons could re-form from free electrons and holes, but the ratio of excitons to free carriers would have to be considerable for excitons to dominate recombination. This ratio can be expected to depend on carrier density, with proportionally fewer excitons than free carriers present as carrier density increases. Dawson et al. compared absorption spectra (in which excitonic domination is expected) to PL spectra for very low carrier densities (8×10^{-8} W/cm² excitation) and originally reported that excitonic recombination dominated PL⁶, and Bimberg, et al. came to a similar conclusion using other techniques.⁷ However, in LEDs carrier densities are much higher, on the order of 10^{18} cm⁻³. The data in Figs. 1 and 2 demonstrate that radiative recombination in GaAs/AlGaAs QWs is dominated by free carriers, not excitons, at high carrier densities.

Figure 1 illustrates room temperature time-resolved PL at the wavelength of the $n=1$ electron to $n=1$ heavy hole transition in a 100 period, 100 Angstrom quantum well, 200 Angstrom Al_{0.26}Ga_{0.74}As barrier MQW structure grown by molecular beam epitaxy (MBE) at AT&T Bell Laboratories. The high-bandgap cladding on the top and bottom of this structure was formed by short-period superlattice layers. The details of the PL measurement system are described in Refs. 8 and 9. The PL decay is exponential with a time constant of approximately 1 nsec over nearly two orders of magnitude in excitation density. (An

excitation of 100 microJ/cm² corresponds roughly to an injected carrier density of 10¹⁸ cm⁻³.) Figure 2 illustrates time-integrated PL energy over the same range of excitation energy density on a logarithmic plot. The slope of this plot is 2.0, indicating a square law dependence. Overall, the PL efficiency of this sample is rather low at room temperature.

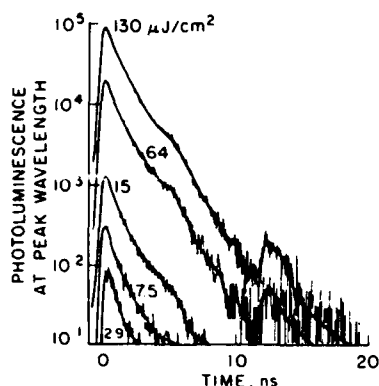


Figure 1. Time-resolved PL (logarithmic scale) of the n=1 electron to n=1 heavy hole transition of an MBE-grown GaAs/Al_{0.26}Ga_{0.74}As MQW structure at room temperature. PL decay is exponential with a time constant of 1 nsec over nearly two orders of magnitude in excitation density. (The second peak and the hump which appears between one and two orders of magnitude down from the main peak are due to optical switch leakage.)

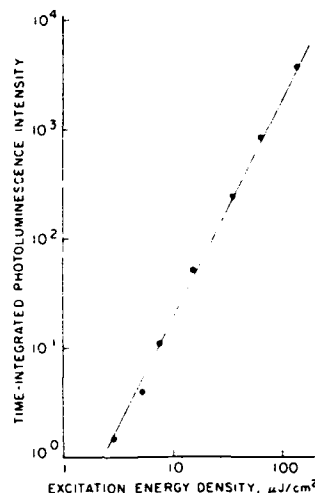


Figure 2. Time-integrated PL as a function of excitation density for the same sample and conditions described in Fig. 1. The slope of the line is 2.0, indicating that time-integrated PL increases as the square of excitation density.

The dynamics of bimolecular radiative recombination in the presence of a linear nonradiative mechanism can be described as

$$\frac{dn}{dt} = -Bnp - r_n n \quad (1)$$

$$\frac{dp}{dt} = -Bnp - r_p p \quad (2)$$

$$\frac{dP}{dt} = Bnp \quad (3)$$

where n and p are the electron and hole densities, respectively, in cm⁻³, B is the bimolecular radiative recombination coefficient in cm³/sec, P is the emitted photon density in cm⁻³, and r_n and r_p are the nonradiative decay rates for electrons and holes in sec⁻¹. When nonradiative decay dominates, the last terms in Eqs. 1 and 2 dominate, so the carrier densities decay exponentially from the optically injected carrier densities. Time-resolved PL then decays as

$$\frac{dP}{dt} = Bn_{inj}^2 e^{-t/T_{nr}} \quad (4)$$

where $T_{nr} = (r_n + r_p)^{-1}$. Hence PL will decay exponentially with the same decay time T_{nr} regardless of excitation density as long as linear nonradiative recombination dominates. Integration with respect to time yields the relation $P = BT_{nr}n_{inj}^2 \propto E_{exc}^2$. Thus time-integrated PL is proportional to the square of excitation density, as shown in Fig. 2.

A similar derivation of excitonic recombination including both radiative and nonradiative components yields exponential decays, but time-integrated PL increases only linearly with excitation density. Therefore spontaneous recombination in GaAs/AlGaAs QWs is due to bimolecular radiative recombination, not excitonic recombination, at room temperature and at carrier densities from mid 10¹⁶ to over 10¹⁸ cm⁻³; recombination in room-temperature LEDs is bimolecular, not excitonic. This conclusion is also supported by the dependence of the PL photon energy as a function of temperature (not shown)¹⁰.

Bimolecular Radiative Recombination Rate

Initially workers had suggested that the excellent performance of GaAs/AlGaAs QW lasers might be due to the difference between the two-dimensional and three-dimensional densities of states³. In three

dimensions, the densities of states for electrons and holes increase parabolically from zero at the bandedge. In two-dimensional structures such as QWs, the density of states for each carrier is zero below the $n=1$ state and constant above it up to the $n=2$ state, where the density of states steps up by this constant value again. Thus more carriers can fit into a narrower energy range next to the bandgap in QW structures. It was hoped that this stronger density of states overlap would lead to a larger bimolecular radiative recombination coefficient, B . The work described here demonstrates that B is not significantly larger in GaAs/AlGaAs quantum wells than in bulk GaAs and is probably smaller.

In the QW structure described above, the domination of nonradiative recombination through the highest carrier densities studied shows that B is considerably less than $5 \times 10^{-10} \text{ cm}^3/\text{sec}$. A similar 40 period QW structure grown by metalorganic chemical vapor deposition (MOCVD) by R.D. Burnham at Xerox with bulk Al_{0.5}Ga_{0.5}As cladding had much slower PL, as shown in Fig. 3. The slow PL decay at even the highest excitation density, corresponding to a carrier density of roughly 10^{18} cm^{-3} , indicates that $B < 7 \times 10^{-11} \text{ cm}^3/\text{sec}$ at excitation densities in the mid- 10^{17} cm^{-3} range. (B has been shown to decrease somewhat with increasing carrier density¹¹, so the value of B at lower carrier densities may be somewhat higher.) Since B is small, then, the excellent performance of GaAs/AlGaAs QW lasers and LEDs is probably due to the funneling of carriers into narrow QWs, creating high carrier concentrations. Then in LEDs bimolecular radiative recombination, which increases as the square of the injected carrier density, competes more effectively with nonradiative mechanisms.

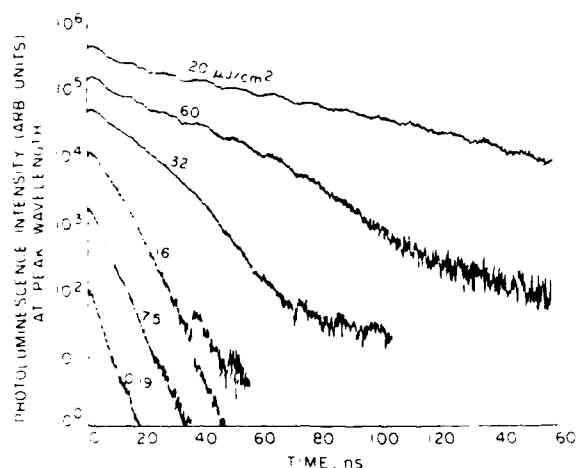


Figure 3. Time-resolved PL of the $n=1$ electron to $n=1$ heavy hole transition of an MOCVD-grown GaAs/Al_{0.23}Ga_{0.77}As MQW structure at room temperature. The relative scaling of the curves taken at different excitation densities is only approximate, but each individual curve follows a consistent scale. At low excitation densities, most of the carriers fall into traps, yielding rapid PL decay from the remaining carriers. At high carrier densities the traps are saturated, yielding a slower decay due to bimolecular recombination.

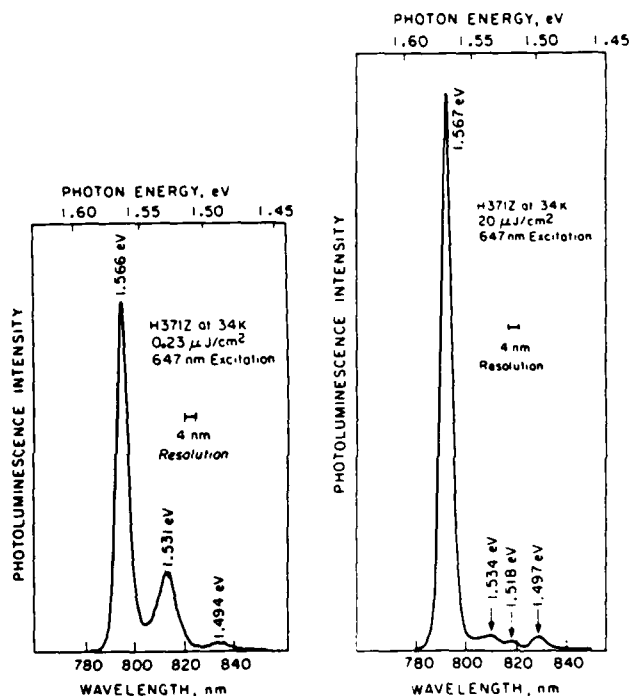


Figure 4. 34 K PL spectra of the sample described in Fig. 3: (a) 0.23 microJ/cm² excitation (b) 20 microJ/cm² excitation. The peak at 1.53 eV is stronger compared to the main peak at lower excitation, characteristic of an impurity peak, not phonon-assisted emission.

Does Phonon-Assisted Stimulated Emission Exist?

A number of GaAs/AlGaAs laser samples optically excited by Holonyak, et al. exhibited lasing at energies below the $n=1$ electron to $n=1$ heavy hole transition. Holonyak, et al. explained this emission as phonon-assisted emission, with the remaining energy going to a longitudinal optical (LO) phonon³. Weisbuch et al. attempted to reproduce this result but could not find any emission at the appropriate wavelength¹². In the spectra shown in Fig. 4, PL is observed at 34 K at an energy 36 meV below the $n=1$ to $n=1$ transition, where 36 meV is the energy of an LO-phonon in GaAs. This emission is demonstrated by time-resolved and time-integrated PL to be luminescence from an acceptor impurity, not phonon-assisted emission. If the observed emission were due to stimulated photon-phonon emission, the effect should be stronger at higher carrier densities. If acceptor-related luminescence were the cause, the effect should be proportionally stronger at low carrier densities, since the limited acceptor density is overwhelmed at high injected carrier densities. Figure 4 shows that the spectral peak at 1.53 eV is proportionally stronger compared to the $n=1$ to $n=1$ transition at lower carrier densities (Fig. 4a) than at higher carrier densities (Fig. 4b), indicating that the low energy peak is due to an acceptor impurity.

Time-resolved PL of this low energy peak should be at least as rapid as time-resolved PL of the main peak if the low energy peak is a longitudinal-optical phonon sideband of the main peak. Figure 5 illustrates that time-resolved PL of the low-energy peak (Fig. 5a) is slower than that of the main peak (Fig. 5b), so the low-energy peak can not be a phonon sideband of the main peak. The slow decay of PL from the low-energy peak is characteristic of acceptor-related PL from both MOCVD and MBE-grown QW samples at low temperatures⁹.

In summary, a PL peak at an energy shift appropriate for phonon-shifted emission was observed at 34 K. This peak is due to acceptor-related PL, not phonon-shifted emission. No peaks at this energy shift were observed at room temperature in these experiments. The explanation for the results of Holonyak et al. is most likely to be associated with experimental geometry as demonstrated by Tarucha, et al.; the spectrum of the emission generated under the pump beam is more strongly absorbed at shorter wavelengths as it exits the sample through the non-pumped region of the waveguide¹³.

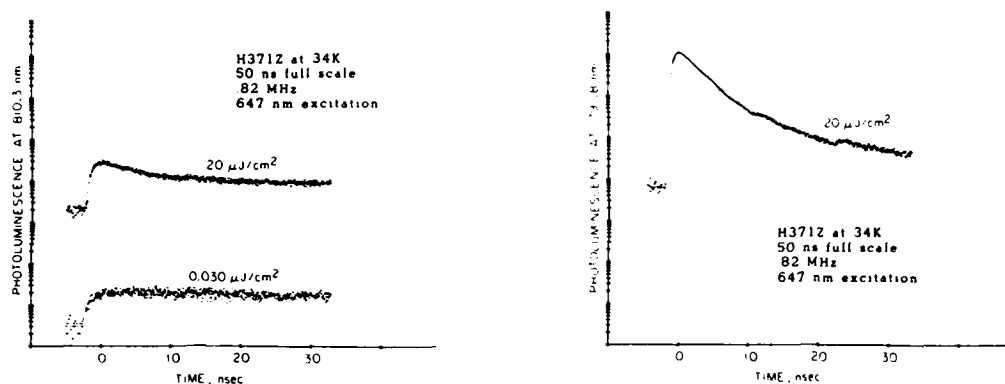


Figure 5. Time-resolved PL of the two spectral peaks described in Fig. 4: (a) peak at 1.531 eV (b) peak at 1.566 eV. PL decay of the low energy peak is due to an acceptor impurity. The low energy peak can not be a phonon sideband of the main peak because it does not decay as rapidly as the main peak.

Trapping

Nonradiative recombination in GaAs/AlGaAs QWs at room temperature depends strongly on material growth. In the QW structures grown by MOCVD, recombination dynamics are dominated by nonradiative recombination through traps at low excitation densities from $\sim 10^{15}$ to $\sim 10^{17}$ cm⁻³. These traps are saturated at higher excitation densities, leaving bimolecular radiative recombination to play a larger role. The expected resulting increase in PL efficiency with increasing excitation density is verified by the time-integrated PL measurements shown in Fig. 6. At room temperature the emitted light increases so strongly with increasing excitation density that it appears the sample may be lasing. However, stimulated emission has a short lifetime, ~ 1 nsec, while Fig. 3 shows that the PL decay time is very long. Therefore, the emission observed at high excitation is due to bimolecular recombination, not stimulated emission.

The behavior of these traps with temperature can also be determined from Fig. 6. At temperatures from liquid helium temperatures to 150 K, time-integrated PL changes little with excitation density. At temperatures above 150 K the traps are activated and become a significant source of nonradiative recombination for low carrier densities. The dynamics of these traps are discussed in more detail in Ref. 14.

The increase in the PL decay rate for a given excitation density over time, as seen most clearly in Fig. 3 for 32 and 60 $\mu\text{J}/\text{cm}^2$ excitation, had not been previously reported in the literature for GaAs/AlGaAs bulk or QW structures to the best of our knowledge. Hence the question arose as to whether this behavior was due to some aspect of the QW structure. Figure 7 illustrates time-resolved PL for a bulk AlGaAs layer of the same Al mole fraction as the barriers of the multiple quantum well structure, with a width similar to that of the QW structure and with similar cladding layers. The QW structure has more interfaces while the bulk AlGaAs has more deep level traps due to Al. Since trapping is more severe in the multiple QW structures (compare to Fig. 3), the trapping is probably associated with the interfaces rather than the bulk AlGaAs.

In the MBE-grown QW structure, the nonradiative mechanism becomes active at a much lower temperature (shown in Ref. 9). Since the nonradiative mechanism could not be saturated at the maximum optical injection level available, its source could not be determined. One likely possibility is deep level traps in the AlGaAs barriers.

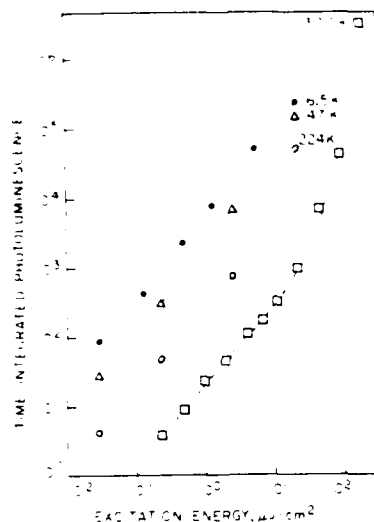


Figure 6. Time-integrated PL as a function of excitation density for the MOCVD MQW structure. At room temperature traps decrease luminescence efficiency at low excitation levels but these traps are saturated at high excitation. They are inactive below approximately 150 K.

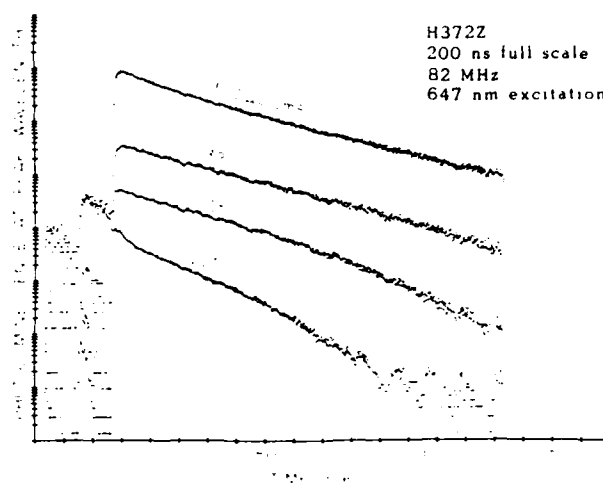


Figure 7. Time-resolved PL of MOCVD bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$. Trapping is not as severe as in the MQW structure shown in Fig. 3 (compare at 7.5 $\mu\text{J}/\text{cm}^2$, for example). Thus the traps are most likely associated with the interfaces in the MQW structure, not the Al in the $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$.

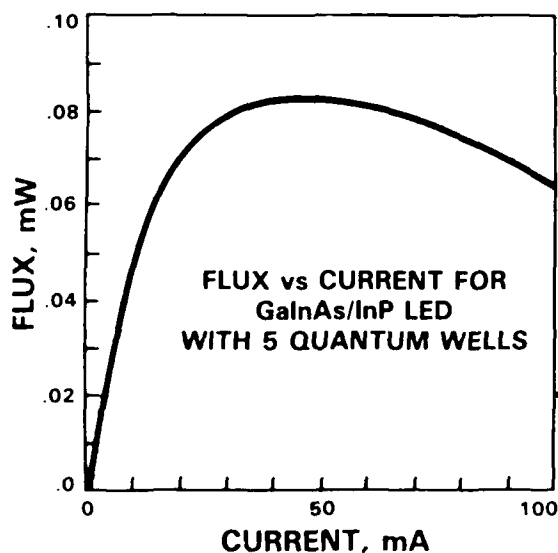


Figure 8. Optical power (approximate scale) as a function of current for an MOCVD $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ 5 QW LED. The saturation is due to increased Auger recombination at high carrier densities.

Auger Recombination

Workers in communications requiring bright sources at 1.3 and 1.55 microns have been interested in applying the successful GaAs/AlGaAs QW designs to longer wavelength operation. However, at these longer wavelengths Auger recombination is much greater than at GaAs wavelengths. Auger recombination

is a three-carrier process in which, for instance, an electron and hole recombine, giving the excess energy to a hole as kinetic energy. Since Auger recombination increases as the cube of the injected carrier density, high carrier densities should be avoided, in opposition to the desired design for GaAs/AlGaAs QW structures. However, the Auger coefficient could be different in QWs compared to bulk GaInAsP. The data of Fig. 8 demonstrate that emission from a 1.3 micron GaInAs/InP quantum well surface-emitting LED saturates and actually decreases with increasing injected current, the result of Auger recombination. These results indicate that the ratio of the Auger recombination coefficient to the bimolecular recombination coefficient is large, roughly $2 \times 10^{19} \text{ cm}^3$ if the assumption is made that no other nonradiative recombination mechanisms are present. Therefore, in order to produce an efficient LED or laser at long wavelengths, it is necessary to avoid high carrier densities by collecting emission more efficiently with a separate confinement heterostructure rather than a surface-emitting structure. To make the most efficient emitter, stimulated emission should be used in a device structure which yields a low threshold carrier density for stimulated emission, such as a separate confinement heterostructure. The stimulated emission rate, not the bimolecular recombination rate, can compete most effectively with Auger recombination at long wavelengths at the carrier densities required for emitters. Even in a separate confinement heterostructure using stimulated emission, the high carrier densities in QWs probably put them at a disadvantage compared to bulk material for long wavelength emitters.

Summary

Conventional PL, electroluminescence, time-resolved PL and excitation-dependent PL were used to study III-V QW structures. Radiative recombination in room temperature GaAs/AlGaAs QW LEDs was found to take place between free carriers, not excitons. The bimolecular radiative recombination coefficient B is quite low, considerably less than $5 \times 10^{10} \text{ cm}^3/\text{sec}$ and probably less than $7 \times 10^{11} \text{ cm}^3/\text{sec}$ for carrier densities on the order of $3 \times 10^{17} \text{ cm}^{-3}$. A low temperature PL peak at an energy appropriate for LO-phonon-shifted emission was found to be due to an acceptor impurity, providing further evidence against LO-phonon-assisted stimulated emission in QWs.

Carrier dynamics were strongly influenced by material-dependent nonradiative mechanisms. Traps in MOCVD-grown GaAs/AlGaAs QW structures appeared to be associated with the interfaces rather than Al-containing material. These traps became saturated at carrier densities used for light-emitting devices. Auger recombination was the primary nonradiative mechanism in 1.3 micron GaInAs/InP surface-emitting LEDs, causing saturation of optical output at a low level. A rough estimation of $A/B \sim 2 \times 10^{19} \text{ cm}^3$ was made.

The success of GaAs/AlGaAs QW light emitting devices is most likely due to the concentration of carriers in narrow QWs in a separate confinement heterostructure, since the bimolecular radiative recombination coefficient is not large enough to yield efficient recombination without high carrier densities. Simply concentrating carriers does not work so well, however, for 1.3 micron and other long wavelength emitters because Auger recombination is so strong and becomes stronger compared to bimolecular radiative recombination with increasing carrier density. If QWs are to prove more advantageous than bulk active regions for any emitters, this advantage will probably be in the case of a low-threshold stimulated emission device.

Acknowledgements

The author is grateful to A.C. Gossard at AT&T Bell Laboratories and R.D. Burnham, now at Amoco Research Center, for growing the GaAs/AlGaAs QW structures. The PL work was done when the author was a student at Stanford University, under an Air Force Office of Scientific Research contract held by A.E. Siegman. AT&T Bell Laboratories and Zonta International provided additional student support. The laser source for PL was loaned by the San Francisco Laser Center, supported by the National Science Foundation. The author wishes to acknowledge useful discussions with M.D. Sturge, now at Dartmouth University; R.C. Miller at AT&T Bell Laboratories and P.M. Fauchet, now at Princeton University.

The MOCVD GaInAs/InP QW LED was grown by K.W. Carey and T.L. Reyes and processed by G. Trott and T. Inouye at Hewlett-Packard Laboratories. G. Trott performed the electroluminescence measurements.

References

1. Nelson, R.J. and Sobers, R.G., Minority-carrier lifetime and internal quantum efficiency of surface-free GaAs, *J. Appl. Phys.*, Vol. 49, p. 6103, 1978.
2. Dingle, R., Confined carrier quantum states in ultrathin semiconductor heterostructures, *Festkörperprobleme (Advances in Solid State Physics)* XV, p. 21, 1975.
3. Holonyak, N., Jr., Kolbas, R.M., Dupuis, R.D. and Dapkus, P.D., Quantum-well heterostructure lasers, *IEEE J. Quantum Electronics*, Vol. QE-16, p. 170, 1980.
4. Hegarty, J. and Sturge, M.D., Exciton holeburning in GaAs/GaAlAs multiquantum wells, *Proc. 17th International Conf. on the Physics of Semiconductors*, Springer-Verlag 1984; and Sturge, M.D., private communication.
5. Chemla, D.S. and Miller, D.A.B., Room temperature excitonic nonlinear-optical effects in semiconductor QW structures, *J. Opt. Soc. Am. B*, Vol. 2, p. 1155, 1985.
6. Dawson, P., Duggan, G., Ralph, H.I. and Woodbridge, K., Free excitons in room-temperature PL of GaAs/Al_xGa_{1-x}As multiple quantum wells, *Phys. Rev. B*, Vol. 28, p. 7381, 1983.

7. Bimberg, D., Christen, J. and Steckenborn, A., Advantages of multiple QWs with abrupt interfaces for light emitting devices, in Two-Dimensional Systems, Heterostructures, and Superlattices, Springer-Verlag 1984.
8. Fouquet, J.E., Siegman, A.E., Burnham, R.D. and Paoli, T.L., Carrier trapping in room temperature, time-resolved photoluminescence of a GaAs/Al_xGa_{1-x}As multiple QW structure grown by metal-organic chemical vapor deposition, Appl. Phys. Lett., Vol. 46, p. 374, 1985.
9. Fouquet, J.E., Recombination dynamics in QW semiconductor structures, Ph.D. dissertation, Stanford University, 1986. Also published by University Microfilms, Ann Arbor, Michigan, 1986.
10. Fouquet, J.E. and Siegman, A.E., Room temperature photoluminescence times in a GaAs/Al_xGa_{1-x}As molecular beam epitaxy multiple quantum well structure, Appl. Phys. Lett., Vol. 46, p. 280, 1985.
11. Su, C.B., Olshansky, R., Manning, J. and Powazinik, W., Carrier dependence of the radiative coefficient in III-V semiconductor light sources, Appl. Phys. Lett., Vol. 44, p. 732, 1984.
12. Weisbuch, C., Miller, R.C., Dingle, R., Gossard, A.C. and Wiegmann, W., Intrinsic radiative recombination from quantum states in GaAs/Al_xGa_{1-x}As multi-quantum well structures, Solid State Comm., Vol. 37, p. 219, 1981.
13. Tarucha, S., Horikoshi, Y. and Okamoto, H., Optical absorption characteristics of GaAs-AlGaAs multi-quantum-well heterostructure waveguides, Japanese J. Appl. Phys., Vol. 22, p. L482, 1983.
14. Fouquet, J.E. and Burnham, R.D., Recombination dynamics in GaAs/Al_xGa_{1-x}As QW structures, IEEE J. Quantum Electronics, Vol. QE-22, p. 1799, 1986.

Recombination Dynamics in GaAs/Al_xGa_{1-x}As Quantum Well Structures

JULIE E. FOUQUET, MEMBER, IEEE, AND ROBERT D. BURNHAM, FELLOW, IEEE

Abstract—Time-resolved and excitation-dependent photoluminescence of GaAs/Al_xGa_{1-x}As quantum well structures reveal that recombination takes place between free carriers, not excitons, at room temperature for carrier densities at and above the mid $\sim 10^{16}$ cm⁻³ level. Other samples show trapping and release of carriers from traps, evidence of dynamic Shockley, Hall, and Read recombination for optically active traps. The traps can be saturated to a large extent. Further studies show that they are associated with interfaces between different materials and that they become active at a temperature around 150 K. Results from all samples indicate that the bimolecular radiative recombination coefficient B for quantum wells is no larger than the value of B for bulk GaAs, and may in fact be smaller. This is one of the first studies of time-resolved luminescence of impurities in quantum well structures. Impurity decays at low temperatures are found to be quite slow. A spectral line which appeared to be longitudinal optical phonon-shifted emission is shown to be due to an acceptor impurity.

INTRODUCTION

RECENT interest in quantum well structures for lasers and other photonic devices has brought about a need to better understand carrier dynamics in these structures. This paper describes experimental studies of recombination dynamics in quantum well structures grown by metalorganic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE). Time-resolved photoluminescence (PL) and excitation density dependent PL are used to determine the roles of bimolecular radiative recombination, excitons, traps, impurities and longitudinal optical (LO)-phonons in quantum well structures.

EXPERIMENT

The MOCVD MQW structure consisted of 40 GaAs (nominally 100 Å thick) wells between 41 Al_{0.23}Ga_{0.77}As barriers (200 Å thick). This MQW active region was sandwiched between two Al_{0.5}Ga_{0.5}As cladding layers transparent to the excitation wavelength; the top cladding layer was 0.9 μm thick while the cladding layer on the substrate side was 4.6 μm thick. The sample was grown on an *n* type (100) oriented GaAs (Si: 3×10^{18} cm⁻³) substrate at 800°C, and the active and cladding regions were not intentionally doped. A Polaron doping profile of

the sample showed a net *n* type carrier concentration of 2×10^{16} cm⁻³ in the active region [1], [2]. The MOCVD SQW structure consisted of a single GaAs well surrounded by two 600 Å wide barriers of the same composition as those of the MQW sample. A 0.6 μm Al_{0.5}Ga_{0.5}As cladding layer was grown over the active region and roughly 3 μm of the same material was grown underneath. The bulk Al_{0.23}Ga_{0.77}As structure was virtually identical to the SQW structure except that no quantum well was grown. These three MOCVD structures were grown at approximately the same time under similar reactor conditions.

The active region of the MBE MQW structure, grown by A. C. Gossard at AT&T Bell Laboratories, nominally consisted of 100 periods of 100 Å GaAs wells and 200 Å Al_{0.26}Ga_{0.74}As barriers on a (100)-oriented Cr-doped GaAs substrate. Short period (13 Å) superlattice buffers of Al_{0.26}Ga_{0.74}As and AlAs having thickness 0.023 μm on either side of the active region functioned as transparent cladding layers and provided a smoothing and cleaning function for the growth surface [3]. Helium-temperature excitation spectra of this sample provided by R. C. Miller at AT&T Bell Laboratories show two different well widths, 93 Å and 100 Å, in roughly equal proportions. The excitation and PL peaks are narrow and appear at the same energy to within the experimental resolution of 1 meV, implying that the interfaces are very flat. The active region and confining layers were not intentionally doped. Pure GaAs grown in this system typically has a net carrier concentration below 5×10^{15} cm⁻³.

Excitation pulses ~ 250 ps long at a typical repetition rate of 820 kHz were provided by a mode-locked krypton ion laser system at 647 nm. These pulses were attenuated, then focused at 50° incidence to illuminate an area $25 \mu\text{m} \times 40 \mu\text{m}$ on the sample for the room temperature experiments. In this experiment, an excitation density of 100 μJ/cm² created a volume carrier density of roughly 10^{18} cm⁻³ in the MQW samples. The resulting PL was passed through a monochromator tuned to the $n = 1$ or other transition for time-resolved measurements. A Hamamatsu R636-P photomultiplier tube and single-photon counting electronics enabled determination of both short and long decay times with wide dynamic range. The FWHM of the detection system response was ≤ 500 ps and the effective time resolution was ≤ 200 ps. The experimental design is described in detail in [1] and [4].

Manuscript received February 21, 1986; revised March 19, 1986. This work was supported by the Air Force Office of Scientific Research, Zonta International, and AT&T Bell Laboratories.

J. E. Fouquet was with the Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305. She is now with Hewlett-Packard Laboratories, Palo Alto, CA 94304.

R. D. Burnham is with Xerox, Palo Alto, CA 94304.
IEEE Log Number 8609658.

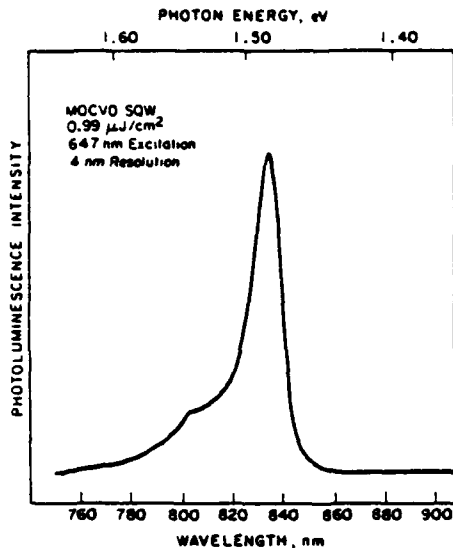


Fig. 1. Time-integrated PL spectrum of MOCVD SQW at room temperature for $1 \mu\text{J}/\text{cm}^2$ excitation (approximately $8 \times 10^{16} \text{ cm}^{-3}$) and 4 nm resolution. The high energy feature is the $n = 1$ electron to $n = 3$ heavy hole transition.

RESULTS FROM MOCVD QUANTUM WELL STRUCTURES Trapping at 300 K

The room-temperature time-integrated PL spectrum of the MOCVD SQW structure is shown in Fig. 1. The steeper low-energy side of the $n = 1$ transition is qualitatively consistent with thermal population of the subbands in the two nonconfined dimensions. Unlike the MBE MQW sample's spectrum, shown in [3], this sample does not show a distinct $n = 1$ light hole transition. The feature near 800 nm in the high-energy tail, shifted about 60 meV from the main peak, can not easily be explained by transitions involving light holes or by luminescence at or above the bulk bandgap energy of the barriers. However, it is likely that this transition is the $n = 1$ electron to $n = 3$ heavy hole transition, using the results of R. C. Miller *et al.* [5]. Regardless of the origin of this feature, the 4 nm spectral resolution used in these experiments excluded its PL from the time-resolved decay measurements of the main PL peak.

The PL spectrum of the MOCVD MQW sample is similar [1] and shows the same high-energy feature but is shifted to longer wavelengths, indicating that the well width in the SQW structure is narrower than the well width in the MQW structure, consistent with the run logs. The ≈ 59 meV shift observed agrees excellently with the 60 meV separation between these two features in R. C. Miller's excitation spectra for 102 Å quantum wells in [5].

Time-resolved PL following the arrival of the excitation pulse, as observed at the peak emission wavelength, is presented in Fig. 2 for the MQW sample at room temperature. At the lowest excitation density $0.19 \mu\text{J}/\text{cm}^2$ the decay is exponential with a decay time of 3 ns. (The secondary pulses visible 36 ns and other multiples of 12 ns after excitation here and in Figs. 3-4, 7, and 9 are due to optical leakage at the Bragg cell. The individual time-re-

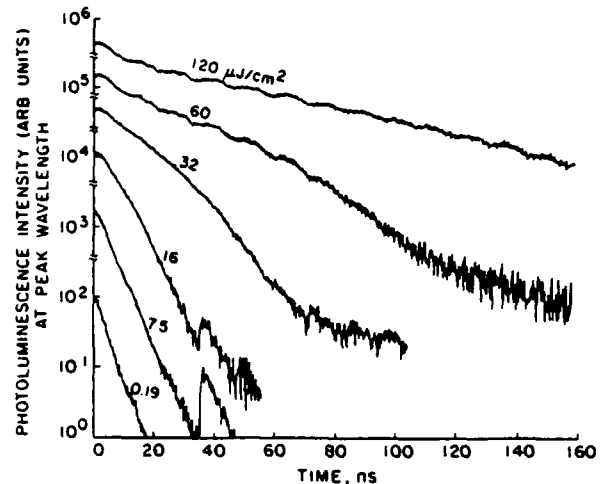


Fig. 2. Time-resolved PL of MOCVD MQW at room temperature for different excitation levels (note logarithmic scale). Approximate carrier densities corresponding to the excitation energy densities are 1×10^{18} , 6×10^{17} , 3×10^{17} , 2×10^{17} , 8×10^{16} , and $2 \times 10^{15} \text{ cm}^{-3}$.

solved decay curves keep to a constant scale, but the different curves are not necessarily normalized to each other.) As the excitation energy density is increased, the initial decay rate slows down. However, for a given excitation density, if one follows the decay to longer times, the rate of the decay eventually increases. In the case of $32 \mu\text{J}/\text{cm}^2$ excitation, a bend can be seen in the PL curve roughly 40 ns after excitation. At $60 \mu\text{J}/\text{cm}^2$ excitation, the speedup occurs later, more than 60 ns after excitation. The high excitation curves decay into noise at what appears to be a high level (three orders of magnitude down from their peaks) because of the attenuation required by the single photon-counting technique. For the excitation density of $120 \mu\text{J}/\text{cm}^2$ the initial decay rate is quite slow, with a decay time on the order of 45 ns, and the delayed speedup has apparently moved outside our observation range. (The 12 ns ripple on the high excitation curves is an artifact of leakage from the laser electronics.)

Trapping behavior has also been observed in the SQW structure grown under similar conditions, as shown in Fig. 3, but in this case the decay rate does not increase over time. Carrier densities generated for given excitation levels in the SQW structure are higher than in the MQW structure because the carriers are funneled into a single well. Assuming all carriers are caught by the single well, effective volume carrier densities will be approximately ten times higher in the SQW than in the wells of the MQW structure. At the lowest excitation, $0.14 \mu\text{J}/\text{cm}^2$, the decay is close to single exponential with a 9 ns decay time. As the excitation density is increased to $0.55 \mu\text{J}/\text{cm}^2$, the decay time lengthens to 19 ns. At the highest excitation density, $108 \mu\text{J}/\text{cm}^2$, the decay is no longer single exponential in character. (The initial peak and rapid decays over the first ≤ 7 ns at the lowest excitation levels in this and the bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample are probably the result of leakage of excitation photons and are not of physical significance.)

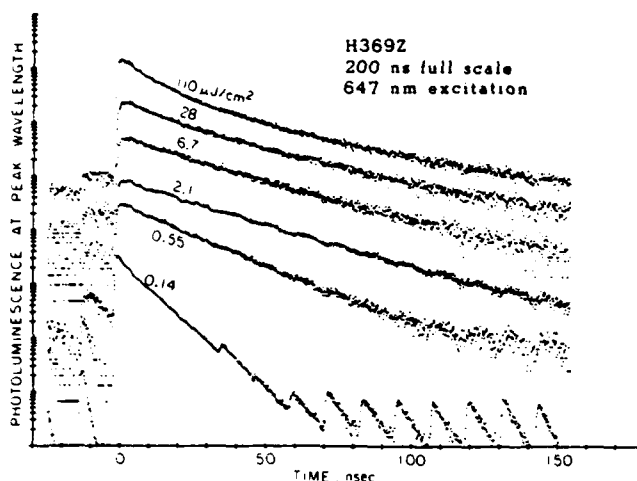


Fig. 3. Time-resolved PL of MOCVD SQW at room temperature for different excitation levels (note logarithmic scale). Approximate carrier densities corresponding to the excitation energy densities are 9×10^{18} , 2×10^{18} , 5×10^{17} , 2×10^{17} , 4×10^{16} , and $1 \times 10^{16} \text{ cm}^{-3}$.

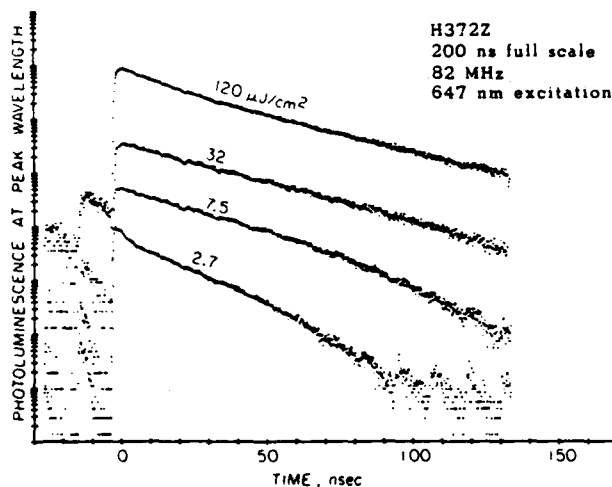


Fig. 4. Time-resolved PL of MOCVD bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ at room temperature for different excitation levels (note logarithmic scale). Approximate carrier densities corresponding to the excitation energy densities are 6×10^{17} , 2×10^{17} , 4×10^{16} , and $1 \times 10^{16} \text{ cm}^{-3}$.

The bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample exhibited trapping and the increase of decay rates, as shown in Fig. 4. At the excitation density of $2.7 \mu\text{J}/\text{cm}^2$ the initial decay time was 19 ns and the later decay time was faster, at 14 ns, while at the highest excitation density, $120 \mu\text{J}/\text{cm}^2$, the decay was not quite single exponential and had a later decay time of 32 ns.

Increases in decay rate like those seen in the MQW and bulk $\text{Al}_x\text{Ga}_{1-x}\text{As}$ samples had not been previously published in the bulk GaAs or GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum well literature when these data were first published [1], to the best of our knowledge. The behavior of time-resolved PL provides a real-time picture of the Shockley, Hall, and Read recombination. The increase in decay rate results in a semilogarithmic PL versus time plot which is concave downward once the decay has been established (Figs. 2 and 4).

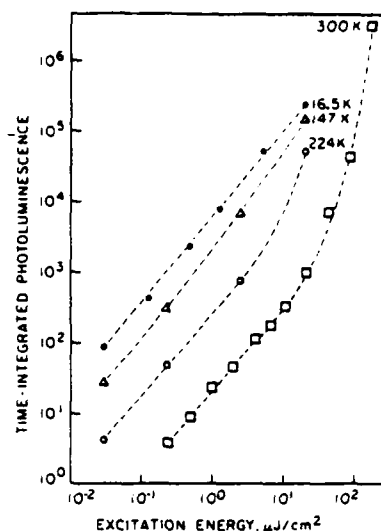


Fig. 5. Time-integrated PL of MOCVD MQW as a function of excitation density at different temperatures.

To rule out experimental artifacts as a cause for this unusual result, a comparative study was made of PL decay from the MQW using excitation pulses at two different repetition rates, 820 kHz and 82 kHz. The later PL decayed slightly more rapidly for 82 kHz excitation than for 820 kHz excitation, which would not be expected for an artifact due to the overlap of the previous decay. In addition, PL just before the arrival of the excitation pulse was observed to be at the noise level due to leakage through the acoustooptic switch used to select pulses, indicating that the previous decay was complete when the next pulse arrived. The more rapid later decay for the 82 kHz excitation repetition rate will be discussed below.

Time-integrated PL of the MQW sample (area under main peak) at room temperature for 647 nm excitation is shown in Fig. 5 as a function of excitation energy density. At room temperature for excitation energy densities below $10 \mu\text{J}/\text{cm}^2$, time-integrated PL increases slightly more strongly than proportionally to excitation energy density. As the excitation density is increased beyond $10 \mu\text{J}/\text{cm}^2$, time-integrated PL increases greatly. In fact, the increase is so great (slope > 2) that it appears the sample is lasing. However, the very slow PL decay times accompanying this increase in PL efficiency (Fig. 2) show that the sample is definitely not lasing, since stimulated emission in bulk GaAs is a rapid process with a characteristic time less than 1 ns [6]. Rather, these results are due to saturation of traps, which will be shown below.

Time-integrated PL as a function of excitation energy density for the SQW sample has a slope of approximately 1.4. The fact that the slope is relatively constant and greater than one over the entire range of excitation densities indicates that at least some bimolecular radiative recombination is present down to the lowest excitation density, $0.32 \mu\text{J}/\text{cm}^2$ [4]. Since this intermediate slope is relatively constant throughout the range of excitation densities, the relative roles of different mechanisms probably do not change significantly. Some nonradiative recombi-

nation is present also. At an excitation density of $110 \mu\text{J}/\text{cm}^2$ the PL efficiency of the SQW sample exceeds that of the MQW sample by a factor of approximately 1.6, even though the MQW structure absorbs over four times as much of the incident radiation as the SQW structure does, due to its thicker and more strongly absorbing active region. Thus, at medium to low excitation densities ($< 10^{17} \text{ cm}^{-3}$) the SQW exhibits considerably higher luminescence efficiency than the MQW structure, so the SQW must not suffer from as much nonradiative recombination.

The observed decrease in decay rate with increasing excitation energy density is characteristic of carriers saturating a limited number of trap sites. At low excitations, most of the carriers fall into traps and a fast decay rate is observed, representing the rapid depletion of the $n = 1$ electron or $n = 1$ heavy hole population. The traps are relatively easier to saturate in the SQW structure than in the MQW structure, however, since a slower decay and higher PL efficiency are obtained in the SQW structure at low excitation levels. Thus it is not unreasonable to expect the effective volume trap density to be lower in the SQW structure than in the MQW structure.

The bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ material showed the decrease in decay rate for higher excitation levels characteristic of trap saturation. In addition, for a given excitation, the sample exhibited an initial slow decay rate, followed by a faster decay rate at later times, similar to the behavior of the MQW structure. The decrease in decay rate with increasing excitation representing trap saturation occurred at excitation densities which were low in the bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ compared to the densities needed to see these effects in the MQW sample (compare Figs. 2 and 4), suggesting that the trap density is higher in the MQW structure than in the bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample. Characteristics of the trap states will be discussed in more detail below.

The MQW and bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ results are quite unusual in that tens of nanoseconds after the establishment of an initial slow exponential decay rate, the decay rate increases. A model to explain these results follows. The time decay of the hole population is determined both by radiative recombination with electrons and by holes falling into trap states (here a hole trap rather than an electron trap is assumed). Bimolecular radiative recombination is given by Bnp where B is the bimolecular radiative recombination coefficient in $\text{cm}^3 \cdot \text{s}^{-1}$ and n and p are the electron and hole populations in cm^{-3} . The rate for holes to fall into the trap states depends on both the density of free holes and the density of available trap state ($S_{\text{max}} - S$) where S_{max} is the total number of trap sites per cm^3 and S is the density of holes in traps, i.e. populated traps.

$$\frac{dp}{dt} = -Bnp - a(S_{\text{max}} - S)p \quad (1)$$

where a is the rate for holes to fall into traps in $\text{cm}^3 \cdot \text{s}^{-1}$. The population of trap state is increased by holes falling in from the $n = 1$ hole quantum levels and decreased by

holes in the traps recombining with free electrons, a process which depends on both the density of holes in the trap state and the density of free electrons.

$$\frac{dS}{dt} = a(S_{\text{max}} - S)p - rSn \quad (2)$$

where r is the rate for holes in the trap state to recombine with electrons in $\text{cm}^3 \cdot \text{s}^{-1}$. Consequently, the electron population is reduced by radiative recombination and by nonradiative recombination with holes in the trap sites

$$\frac{dn}{dt} = -Bnp - rSn. \quad (3)$$

Time-resolved PL results from the radiative recombination process only and is given by

$$\frac{dP}{dt} = Bnp \quad (4)$$

where P is the density of emitted photons in cm^{-3} . The coefficients B , a , and r are temperature dependent.

At low excitation densities, PL decay in the MQW sample is dominated by the decay of carriers into trap states, the last term in (1). Carriers which fall into these traps do not contribute to radiative recombination at the $n = 1$ transition energy. Thus the time-integrated PL efficiency out of the sample is low. The number of traps is limited, however, and at high excitation densities it is possible to fill most of the traps. The highest excitation density in this experiment, $120 \mu\text{J}/\text{cm}^2$, corresponds to an injected carrier density of over $1 \times 10^{18} \text{ cm}^{-3}$. At this excitation level, many carriers are left in the $n = 1$ levels in the conduction and valence band and are able to radiatively recombine at the slower radiative rate. Since the initial carrier densities are directly proportional to the optical excitation density, the radiative rate increases at high excitation densities, further enhancing radiative recombination over the competing nonradiative mechanism(s). Thus, the large increase in room temperature time-integrated PL efficiency at high excitation densities in Fig. 5 is explained. (If we had been able to completely saturate the trap, the 300 K data in Fig. 5 would have leveled off to unity slope representing the radiative limit, i.e., *light out* = *light in*.)

The increase in PL decay rate some time after excitation may be explained by a finite lifetime for carriers in the trap states. Carriers in the traps eventually fall into other levels within the bandgap or directly recombine. When the trap has lost its carrier via the second term in (2) it is free to accept another. As the populations in the $n = 1$ electron and hole levels are depleted by radiative and nonradiative recombination, the trap plays a proportionally larger role. For fast decay times into the traps and fast release times from the traps, the PL decay rate will increase during the course of a decay, as observed in Fig. 2 and Fig. 4.

This model is basic. A more sophisticated model would allow carriers to proceed from one trap state to another before recombining, would allow carriers to ionize out of

the trap back into the $n = 1$ levels, and would take background carrier densities into account in a more sophisticated manner. As it stands, however, the model has four free parameters, which are enough to make a solution difficult.

A rough numerical solution of these rate equations was carried out using a variable step size which was calculated so as to keep the percentage change in the carrier populations small for each step. With the right parameters, the numerical model yielded the right characteristics: a rapid decay for low excitation, a slow decay for high excitation, and an increase in decay rate as time proceeded.

For the 82 kHz excitation rate, more traps are able to empty between excitation pulses. Hence a larger fraction of carriers created by the subsequent excitation pulse fall into traps and the PL decays more rapidly than for the 820 kHz excitation rate. This behavior was also demonstrated using the numerical rate equation model.

The time-resolved PL results of Figs. 2 and 4 can not be explained by the dependence of B on carrier density as studied by Su *et al.* in bulk GaAs [7]. The carrier densities in this experiment were too low to significantly change B , and even if the carrier density dependence of B were different in quantum wells, time-resolved PL due to this effect bends the logarithmic plots of time-resolved PL downwards soon after excitation because B changes only at high excitation, unlike the data here in which the bends occur much later.

The time-resolved PL of the SQW structure shown in Fig. 3 has an initial rapid, concave upwards decay at high excitation which can be explained by either bimolecular recombination, diffusion or Auger recombination. These high carrier density effects are more likely to appear in the SQW structure because the funnel design creates higher carrier densities. The fact that the slope of the logarithmic time-integrated PL versus excitation plot is constant rather than decreases at high excitation densities makes bimolecular recombination more likely than diffusion and Auger recombination since the latter two processes are nonradiative and would decrease the luminescence efficiency. The absence of increases in decay rate in the SQW time-resolved PL can be understood as follows. Because of funnelling, n and p will be large, making Bnp large compared to its value for the MQW and bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ structures. Assuming that the traps are indeed interface traps, as will be shown below, and that the SQW structure has fewer of them per unit volume than the MQW structure, the S_{max} in cm^{-3} will be smaller for the SQW than the MQW structure. Thus the trapping terms are smaller in the SQW structure and the bends are not seen in the PL decay. This argument is also favored by the higher PL efficiency of the SQW structure.

Source of Traps

The precise source of the trap sites is not known, but a few deductions limit the range of possibilities. Bulk $\text{AlGa}_{1-x}\text{As}$ often suffers from traps, so the barrier material may be the source of the traps. However, the larger

role of interfaces in quantum well structures could lead to the observation of interface traps which have not been seen in bulk material.

The bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample (Fig. 4) investigation helped to clarify the question of trap origin. If the traps had been in the bulk ternary material, trapping would have been worse in this bulk sample than in the MQW structure. Because the trapping problem was not as serious in the bulk sample, whose active region has just two interfaces located at the cladding layers, as it was in the MQW sample, which has 81 interfaces, the traps are probably at the interfaces. (Compare Fig. 2 to Fig. 4; the PL decay rate increased at lower excitation levels in the bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample than in the MQW sample, indicating that the traps are easier to saturate in the bulk $\text{Al}_{0.23}\text{Ga}_{0.77}\text{As}$ sample than in the MQW sample.)

Value of B

The magnitude of the bimolecular radiative recombination coefficient B can not be determined directly from time-resolved PL because the nonradiative trap mechanism dominates the decay. However, indirect determinations of B can be made in two different ways. First, the long decay time at high excitation can be used to put an upper limit on the value of B . (In this case, the initial rapid decay will be ignored because it may be due to diffusion.) The main exponential decay is most likely due to falling into trap states, i.e., nonradiative recombination; neither radiative recombination nor diffusion look exponential. The 45 ns decay time at the highest excitation level implies a decay rate of approximately $2.2 \times 10^7 \text{ s}^{-1}$. The radiative recombination rate Bn must be less than or equal to this rate. The maximum value of n is taken to be $3 \times 10^{17} \text{ cm}^{-3}$ as an approximation of the magnitude of n once the initial rapid decay is complete. Then we can write $B \leq 7 \times 10^{-11}$.

The other way to determine the value of B lies in numerical solution of the rate equations above. These rate equations are quite sensitive to the value of B . A rough solution neglecting repetitive excitation to reach steady state was carried out. To generate decays as slow as the data, it is necessary to use a small value of B , on the order of a few times 10^{-11} . Thus, these two approaches yield consistent predictions of B , and B is small. Due to the necessarily indirect manner in which the evaluations of B were made and to the difficulties of precisely determining carrier densities, these results should be considered more as order-of-magnitude values than precise measurements.

Slow Impurity PL at Low Temperatures in MOCVD MQW

Fig. 6 shows the 1 nm resolution PL spectrum at 16.5 K of the MQW sample for $0.030 \mu\text{J}/\text{cm}^2$ excitation. The main peak is accompanied by a series of peaks, centered at 1.548 eV, possibly due to a series of acceptor impurities. Also, a shoulder is present on the low energy side of the main peak which is probably due to donor luminescence, using the results of Bastard [8]. The large acceptor

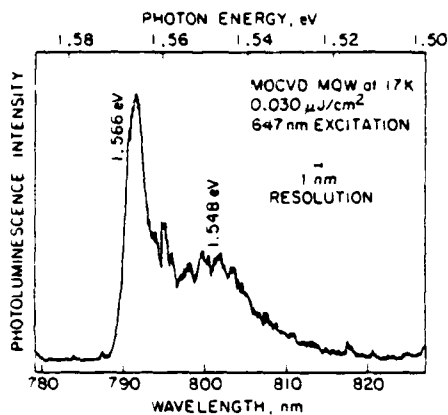


Fig. 5. Time-integrated PL spectrum of MOCVD MQW at 17 K for 0.03 $\mu\text{J}/\text{cm}^2$ excitation (approximately $3 \times 10^{14} \text{ cm}^{-3}$) and 1 nm resolution. The structure around 1.548 eV is due to electron to acceptor luminescence.

PL peaks in this n type sample suggest a relatively large impurity density. At higher excitation levels the PL spectrum (not shown) changes somewhat. The peak of the impurity series around 1.548 eV moves towards higher photon energies, for instance to approximately 1.558 eV for 0.48 $\mu\text{J}/\text{cm}^2$ excitation. This series becomes less distinct as the excitation density is increased to higher levels, turning into a shoulder on the side of the main peak.

The shift of the acceptor peaks to higher energy with increasing excitation density is reminiscent of, but larger than, the results of Lambert *et al.* who also observed a greater shift in acceptor ($e - A^0$) luminescence to shorter wavelengths than in donor ($D^0 - h$) luminescence with increasing excitation. The magnitude of the shift they observed was approximately 3 meV for an increase in excitation density from 10 mW/cm² to 1 kW/cm² [9]. By contrast, in the results presented here the acceptor-related luminescence peak wavelength shifted 10 meV higher in energy between just 0.030 and 0.48 $\mu\text{J}/\text{cm}^2$ excitation (these pulsed densities fall within Lambert *et al.*'s range when converted to CW values).

Shifts of impurity peaks to higher photon energies with increasing excitation density have been observed for donor to acceptor transitions ($D^0 - A^0$) in bulk III-V materials as reviewed in [10]. Donor-acceptor transitions are thought by some workers to be rare in quantum wells, but other workers have identified peaks at an energy shift of approximately 47 meV as donor-acceptor transitions [9]. The total energy shifts of the peaks observed here appear to be small for donor-acceptor shifts. However, the more recent results of Brum, Bastard, and Guillemot indicate that the large impurity density in this sample may reduce the impurity binding energies [11] so that the luminescence could be coming from acceptors distributed uniformly through the wells. Another argument against donor-acceptor recombination is that the peak spacing for typical pair separations observed in donor-acceptor recombination is smaller than that observed here. Thus, the separate peaks here are more likely to be due to an acceptor series than to donor-acceptor luminescence.

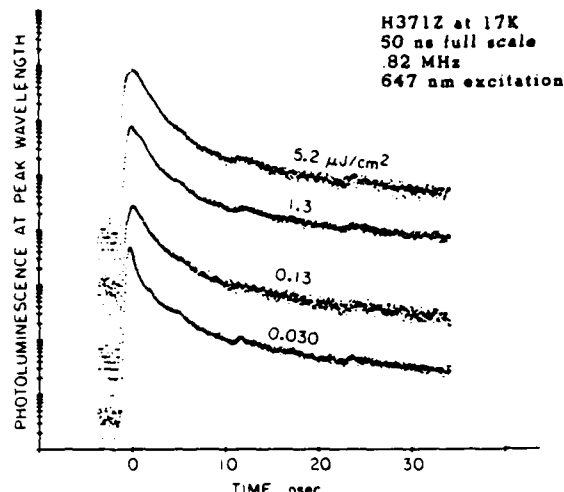


Fig. 7. Time-resolved PL of MOCVD MQW at 17 K for different excitation levels. Approximate carrier densities corresponding to the excitation energy densities are 5×10^{16} , 1×10^{16} , 1×10^{15} , and $3 \times 10^{14} \text{ cm}^{-3}$.

Time-resolved PL of the MQW structure at 16.5 K at the wavelength of the main peak is shown in Fig. 7 (note time scale). The decays are faster than at room temperature and highly nonexponential. Time-resolved PL of the donor feature at 1.559 eV is quite a bit slower than that of the main peak. The acceptor luminescence around 800 nm also decays more slowly than the main peak.

MQW and SQW luminescence at 16.5 K are probably predominantly excitonic, as shown by the weak dependence of decay time on excitation in Fig. 7 and by comparison of the photon energy of the peak at low temperatures to the peak at room temperature (not shown). At the highest excitation levels the exciton population may become somewhat saturated, leading to more bimolecular recombination, which is slower. The decays are not single exponential. Carriers falling into impurity or trap states are partially responsible for the decays observed here. In a quantum well, the impurity wave functions depend on the location of the impurity within the well, so the rate to bind to a particular species of impurity differs between individual atoms of that species. Thus, one impurity species could appear like a series of different impurities in bulk material. Another possible explanation for the data is trap saturation. If a small number of trap or impurity states are available, most of the carriers may fall into them at low excitation. At high excitation the trap population is saturated. It is not possible to determine exactly which mechanisms are responsible for the details of the behavior observed here with the limited information available. The SQW structure did not exhibit nearly as much impurity-related PL as the MQW.

Time-integrated photoluminescence of the main peak had a slope on a logarithmic plot greater than one throughout the range of excitation densities, indicating that not all the carriers were recombining radiatively. The corresponding data for the acceptor feature around 1.55 eV showed a slope greater than one at low excitation levels which may be due to pumping by the main line, feed-

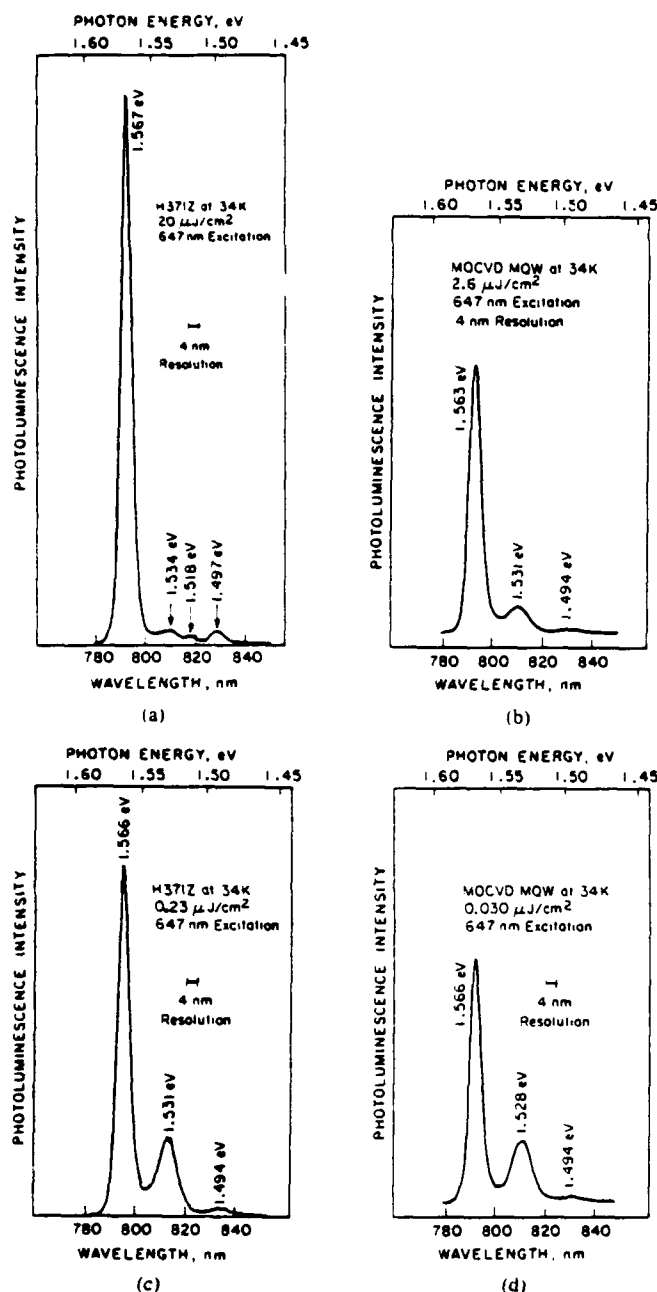


Fig. 8. Time-integrated PL spectra of MOCVD MQW at 34 K for four different excitation levels corresponding to 2×10^{17} , 3×10^{16} , 2×10^{15} , and $3 \times 10^{14} \text{ cm}^{-2}$. The impurity peak around 1.531 eV is readily apparent at low excitation, but becomes much weaker compared to the main PL peak at high excitation.

ing carriers to lower-lying impurity states, or both. This feature saturates at higher excitation densities, as expected for an impurity.

36 meV-Shifted Emission is Not Phonon-Assisted Emission

As the temperature is increased to 34 K the character of the spectrum changes, as can be seen by comparing Fig. 8(d) to Fig. 6. The main peak is still present at 1.566 eV. The feature at 1.494 eV is an acceptor transition in the bulk GaAs buffer layer and/or substrate. A peak of con-

siderable magnitude arises at 1.528 eV for $0.03 \mu\text{J}/\text{cm}^2$ excitation. An increase in excitation density shifts this peak to shorter wavelengths, as can be seen in Fig. 8. The difference in energy between this peak and the main peak is approximately the energy of an LO-phonon in GaAs, 36 meV. As the excitation level increases, this peak becomes relatively weaker compared to the main peak.

Longitudinal optical (LO)-phonon replicas of various emission lines are often observed in bulk III-V compounds, but their existence in quantum wells has been disputed. Holonyak's group at the University of Illinois has reported lasing in quantum wells on a transition very roughly 36 meV below the $n = 1$ heavy hole transition, and has attributed it to stimulated phonon emission [12]. Other researchers have not been able to reproduce this result [13]–[15], and in fact could not find a spectral line at the appropriate energy.

Time-resolved PL of both the peak approximately 36 meV lower in energy and the main peak, shown in Fig. 9, demonstrates that the main peak decays *much more rapidly* than the low-energy peak. In contrast, to a first approximation one would expect a phonon sideband to decay at least as rapidly as the peak of which it is a sideband. Thus these data indicate that the low energy peak is not due to phonon-shifted emission. The low-energy peak also has a slower decay than the main peak at 58 K. Slow PL decays at low temperatures are correlated with impurity PL in the MBE sample (discussed below) and at other temperatures in the MOCVD samples, suggesting that the feature here is an impurity.

Time-integrated PL also indicates that this feature is not due to LO-phonon emission. Comparison of the spectra at different excitation densities in Fig. 8 shows that the 1.531 eV peak plays a proportionately larger role at *low* excitation, suggesting this peak is due to an impurity because the limited number of impurity atoms can be saturated. A stimulated process should be stronger compared to the main peak at high excitation. In addition, the peak shifts upwards in energy by 6 meV from the lowest excitation level ($0.030 \mu\text{J}/\text{cm}^2$) to the highest ($20 \mu\text{J}/\text{cm}^2$). As mentioned above, shifts of this sort were observed for acceptor-related luminescence, further strengthening the argument that this feature is acceptor-related. Impurity identification is not trivial in quantum wells, but using the results of Bastard [8] it appears that the impurity is a shallow acceptor, most likely carbon.

The conclusion that this peak at 34 K results from an acceptor impurity does not necessarily mean that the features Holonyak's group observed at room temperature were caused solely by this mechanism. The experimental setup employed by that group [12] usually optically excited only part of the length of the active region of the sample. The rest of the sample then acted as an absorber for lasing in the lengthwise direction, removing spectral components above the $n = 1$ transition energy and leaving emission at longer wavelengths intact. As a general rule, Holonyak's lasers tended to lase more at longer wavelengths in the lengthwise direction than in the crosswise

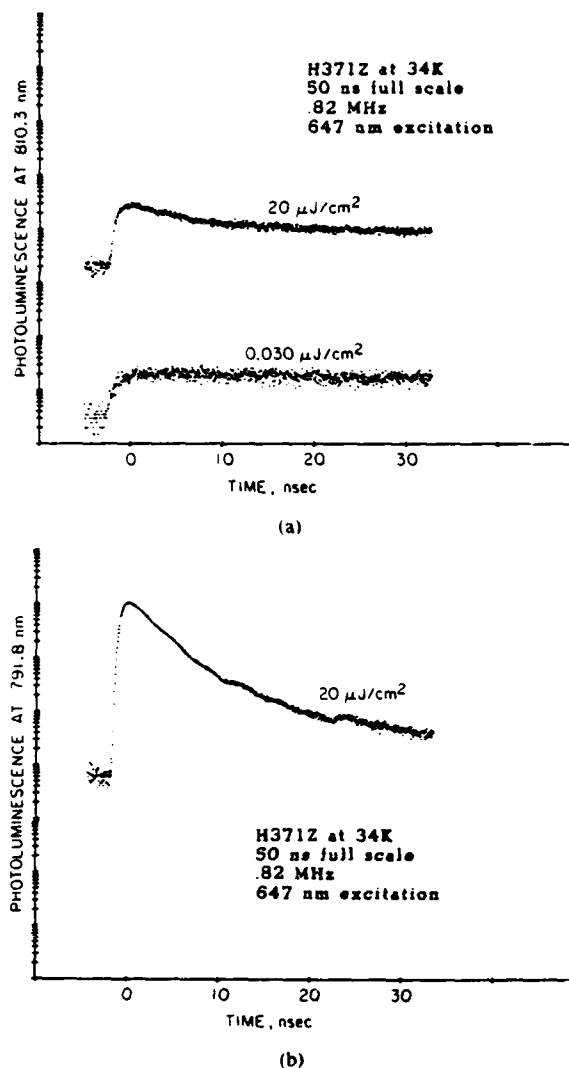


Fig. 9. Time-resolved PL of MOCVD MQW at 34 K. (a) Long wavelength impurity peak. (b) Main peak. $20 \mu\text{J}/\text{cm}^2$ corresponds to approximately $2 \times 10^{17} \text{ cm}^{-3}$ in this sample.

direction, which was pumped along a greater fraction of its distance [12].

Tarucha *et al.* used a geometry similar to that of Holonyak *et al.*, but also varied the distance l between the excitation spot and the edge of the sample. PL out of the surface of the sample closely represented the true spontaneous emission spectrum and did not change with l . Room temperature PL out of the end of the sample depended strongly on l , and when the PL had to travel through a long distance of passive waveguide ($l \sim 45 \mu\text{m}$) another spectral peak appeared at longer wavelengths as a result of selective absorption of shorter wavelength emission. The energy shift depended strongly on the particular wafer and on temperature [16].

This absorption loss resulting from Holonyak *et al.*'s experimental geometry will certainly favor otherwise improbable lasing at longer wavelengths and explains the wide range of shifts observed by Holonyak *et al.* (the range was so great, in fact, that it made discrete LO-phonon shift seem unlikely as an explanation). Still, some

sort of gain mechanism is required to achieve laser action at the longer wavelengths. Acceptor transitions and band tailing, bandgap renormalization, and wells of width considerably greater than average are possible explanations.

Fig. 5 illustrates time-integrated PL as a function of excitation energy density for the MQW structure at a number of different temperatures. The PL efficiency at low excitations is much higher at low temperatures than at room temperature. PL efficiency does not decrease significantly between the lowest temperatures and 150 K. An Arrhenius plot (not shown) confirms that the trap is indeed activated at ~ 150 K. The slope of the data at 16.5 K is roughly 1.2, indicating that recombination is not entirely radiative. The behavior of time-integrated PL at 224 K exhibits the same qualitative trap saturation characteristic as at room temperature.

As the temperature increases, more of the recombination becomes bimolecular as excitons ionize, and bimolecular recombination becomes slower. These effects explain the general increase in decay time from the lowest temperatures to room temperature.

RESULTS FROM MBE MQW

Free Carrier Recombination at 300 K

Room temperature recombination was shown to take place between free carriers, not excitons, in [3] and [4]. These results will be summarized here. Time-resolved PL measurements at room temperature of the $n = 1$ heavy hole transition in the MBE MQW structure showed a single-exponential decay with $\tau \approx 1$ ns at excitation densities from 3 to $130 \mu\text{J}/\text{cm}^2$. Time-integrated PL increases as the square of excitation density. These data can be explained by only bimolecular recombination in the nonradiative limit, *not* by recombination of free excitons or monomolecular recombination. In addition, a comparison of photon energy versus temperature with bulk GaAs behavior shows that the excitons have a binding energy of ≈ 8 meV and ionize at a temperature < 50 K [3], [4]. This shift of photon energy to lower values at low temperatures must be due to formation of excitons rather than localization in wide areas of wells with irregular interfaces because R. C. Miller's low-temperature excitation and PL spectra show that the interfaces are very flat.

Bound excitons can not be responsible for the room temperature dynamics in this sample. No bound excitons were present in the helium temperature photoluminescence spectrum taken by R. C. Miller, and since bound excitons disappear rapidly with increasing temperature [17], none should be expected at room temperature.

Finally, free exciton radiative recombination should be rapid ($\gamma_{\text{rad}} \sim 1$ to $2 \times 10^9 \text{ s}^{-1}$) at low temperatures and the exciton radiative decay rate should be similar at room temperature since the mechanism is not directly temperature dependent [18]. Therefore, PL efficiency would be much higher than it is if many excitons were playing a significant role in PL.

The excitation densities for which this conclusion, that room temperature recombination takes place between free

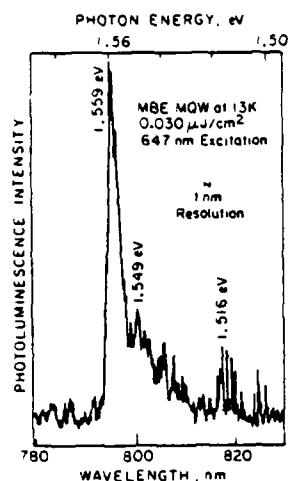


Fig. 10. Time-integrated PL spectrum of MBE MQW at 13 K for $0.03 \mu\text{J}/\text{cm}^2$ excitation (approximately $3 \times 10^{14} \text{ cm}^{-3}$) and 1 nm resolution. The main peak is due to free excitons while the peak at 1.549 eV is due to electron to acceptor luminescence. The peak around 1.516 eV emanates from the bulk GaAs substrate.

carriers, holds correspond to equivalent volume carrier densities in the mid 10^{16} cm^{-3} range and upward. Thus, for light emitting device applications, which require high carrier densities, the recombination mechanism will be free carrier recombination, not excitonic recombination. These results do not specify whether or not recombination at low excitation densities (10^{15} cm^{-3} and lower) takes place between free carriers or excitons. Work by other groups at low excitation densities suggests that recombination at low densities may be excitonic [19]–[20].

Value of B

The persistence of the 1 ns decay time and the slope of 2.0 on the time-integrated PL plot indicate that the non-radiative mechanism has not been saturated at the highest excitation densities, above $100 \mu\text{J}/\text{cm}^2$, which generate carrier densities greater than 10^{18} cm^{-3} . This information indicates that B is considerably less than $5 \times 10^{-10} \text{ cm}^3 \cdot \text{s}^{-1}$ at room temperature [3], [4]. The nonradiative mechanism is discussed in [3] and [4].

Slow Impurity-Related PL at Low Temperatures in MBE MQW

The low temperature spectrum at $0.030 \mu\text{J}/\text{cm}^2$, Fig. 10, is simple compared to that of the MOCVD-grown MQW (Fig. 6) because the MBE MQW contains a considerably smaller number of electrically active impurities. The main peak shifts approximately two meV toward higher energies for higher excitation levels, up to $20 \mu\text{J}/\text{cm}^2$. This shift is probably due to partial saturation of the exciton population at high excitation levels, leading to more free carrier recombination. In fact, it is notable that this shift is relatively small, i.e., the exciton population is not entirely saturated at the highest excitation densities. The free heavy hole exciton peak around 1.559 eV is accompanied by two peaks, at 1.549 eV and 1.531 eV. Assuming an exciton binding energy of 8 meV in these 100

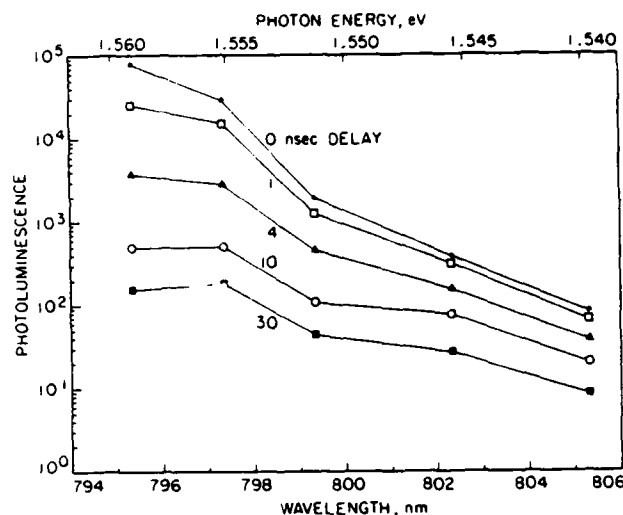


Fig. 11. Instantaneous PL spectrum of MBE MQW as a function of time at 15 K and 1 nm resolution (note logarithmic scale). The free exciton peak around 1.559 eV decays much more rapidly than acceptor-related PL at longer wavelengths. The excitation energy density of $20 \mu\text{J}/\text{cm}^2$ corresponds to approximately $2 \times 10^{17} \text{ cm}^{-3}$.

Å quantum wells from [3], these peaks are then shifted 17 meV and 35 meV, respectively, from the free $n = 1$ heavy hole transition. Both shifts are too large for donor-hole transitions in quantum wells according to the work of Bastard [8], so they most likely involve acceptors.

Using the results of Bastard, the binding energy for carbon acceptors in the center of a 100 Å well is 34 meV, while the binding energy at the interfaces is 13 meV. Thus the 17 meV shift of the 1.549 eV peak is a reasonable binding energy for carbon atoms near the interfaces. R. C. Miller has observed a 20 meV acceptor binding energy in 100 Å quantum wells and has attributed it to carbon near the interfaces in other samples from this source [21]. The peak at 1.531 eV in Fig. 10 could also be due to carbon at the center of the wells but may be another acceptor. Multiple-acceptor species are probably present, and donors may be luminescing at an energy just below the main peak. Time-integrated PL at higher excitation densities does not in general exhibit as much impurity recombination relative to the main PL peak because the injected carriers overwhelm the impurities.

Time-resolved PL of the main spectral peak (not shown) is similar to that of the MOCVD MQW shown in Fig. 7, but is somewhat more rapid. Similar mechanisms are responsible, but more nonradiative recombination is present in the MBE sample at low excitation densities, making those decays more rapid than the corresponding MOCVD results.

Time-resolved PL of the different spectral peaks varied greatly. Fig. 11 illustrates this difference as rough spectra at different times for an excitation density of $20 \mu\text{J}/\text{cm}^2$ at 15 K, shown on a logarithmic scale. These data were obtained by studying time-resolved PL at a number of different wavelengths while keeping all system parameters constant, so that all the decay curves would be normalized to the same value. Then the decay data at a particular time

were extracted from the curves taken at different wavelengths. The main PL peak around 1.558 eV decays much more rapidly than the long wavelength features. For instance, 10 ns after excitation, the main peak has decreased by over two orders of magnitude while PL at 1.549 meV has decreased by an amount closer to one order of magnitude. At long delays the free exciton peak is comparable in magnitude to the acceptor peak, even though the free exciton peak was much stronger initially. Thus acceptor PL decays much more slowly than free exciton PL. Donor-hole PL is slightly slower than PL of the main peak.

This work, along with the MOCVD work described above, comprises one of the first studies of impurity luminescence in quantum wells, to the best of our knowledge. Bimberg *et al.* have published time-resolved cathodoluminescence spectra for MQW's. They also observe slower acceptor-related luminescence for 113 Å MQW's, but not for 55 Å Be-doped MQW's [22]. The fact that the impurity luminescence decays more slowly than the free exciton luminescence is not surprising. The matrix element for impurity transitions at low impurity densities and high excitation levels in bulk is generally weaker than that for free carrier transitions. Excitonic recombination is even stronger than free carrier recombination because of the better spatial overlap of the electron and hole wave functions. Thus, the free exciton recombination rate should be higher than the impurity recombination rate. Donor-acceptor recombination can be even slower than electron-acceptor recombination because of the small spatial overlap of wavefunctions localized to separate donors and acceptors. It is possible that donor-acceptor luminescence contributes to the progressively slower decays at longer wavelengths, but most of the luminescence appears at wavelengths shorter than those appropriate for donor-acceptor recombination. The increase in PL decay time with increasing wavelength suggests that "feeding" processes may be influencing the observed PL decays. That is, excitons ionize and the holes fall into shallow impurities and from the shallow impurities into deeper impurity species, or to impurities of the same species which are closer to the center of the well, so that PL moves out to longer wavelengths. These results are not due to trapping in isoelectronic traps formed by width fluctuations of the wells because R. C. Miller's low-temperature studies showed that the well widths are very uniform.

Diffusion is probably not playing a major role in the dynamics observed here. The dynamics around 77 K where mobilities are the highest, are single exponential. Diffusion dynamics in two dimensions are not single exponential. At no temperature did the PL dynamics exhibit the characteristics of simple diffusive decay.

PL intensity increases strongly as temperature decreases. At 12.5 K, the slope of time-integrated PL of the main peak versus excitation density on a logarithmic plot (not shown) is roughly 1.4, indicating that at least some nonradiative recombination is present, even at this low temperature. At 22 K, the slope of this peak has in-

creased to roughly 1.8, approaching the room temperature value of 2.0, and the PL efficiency at low excitation densities has decreased significantly, indicating that the non-radiative mechanism is playing a larger role. This behavior differs markedly from that of the MOCVD MQW structure, in which little change from the low temperature PL was seen until the temperature was raised as high as 150 K. The acceptor-related photoluminescence peak around 1.536 eV can be spectrally resolved well enough at 22 K to determine the slope of time-integrated photoluminescence on a logarithmic plot. This slope starts off with a value well over 1.0 but less than 1.8, which may be due to pumping by the main line, feeding carriers to lower lying impurity states, or both. At higher excitation levels this transition saturates.

CONCLUSION

Room temperature time-resolved PL in the MOCVD MQW structure depends strongly on excitation energy density. On a gross scale, the short (3 ns) lifetime at low excitation which lengthens to a lifetime an order of magnitude longer at high excitation is typical of the saturation of carrier trap states. The observed increase in time-integrated PL efficiency at high excitation is also characteristic of saturation of carrier traps. PL decay rate at a given excitation level is observed to increase tens of nanoseconds after excitation. This behavior may be explained by the emptying and subsequent refilling of these trap states, which must have limited holding times, and is an example of a dynamic Shockley, Hall, and Read recombination. These traps are probably interface traps rather than bulk traps and are not active at temperatures below approximately 150 K. Two rough determinations of the bimolecular radiative recombination coefficient B yield consistent results in the mid $10^{-11} \text{ cm}^3 \cdot \text{s}^{-1}$ region at 300 K.

Low temperature time-resolved PL from MOCVD quantum well structures is not single exponential in character. Excitonic recombination alone would yield an exponential PL decay, but carriers also fall into a large number of impurity states, making the decays more complex. Recombination at donor- and acceptor-shifted energies is slower than at the main peak. Acceptor-related luminescence shifts to higher photon energies at higher excitation levels. A spectral feature at an energy appropriate for an LO-phonon shifted transition was observed in the MQW structure. However, both time-resolved and time-integrated PL studies of the feature show that it is due to an acceptor impurity, probably carbon at the interfaces between the wells and the barriers.

Room temperature PL in the MBE MQW sample is dominated by free carrier recombination rather than excitonic recombination at injected carrier densities over mid 10^{16} cm^{-3} . B is considerably less than $5 \times 10^{-10} \text{ cm}^3 \cdot \text{s}^{-1}$ at room temperature.

Low temperature PL spectra of the MBE MQW show relatively few electrically active impurities. Time-resolved PL of the main spectral peak shows multiple de-

cays including the heavy hole exciton decay and a range of decays corresponding to carriers falling into a variety of impurity states. At the highest excitations, the exciton population starts to saturate and slower free carrier bimolecular recombination appears. This work and the work on the MOCVD quantum well structures comprise one of the first published studies of time-resolved PL of impurities in quantum wells. PL involving acceptors is considerably slower than PL of the main spectral peak. PL involving donors appears to be slower than PL of the main peak, but the effect is not as marked as for acceptors. The monotonic increase in decay time with increasing wavelength suggests that "feeding" of the lower lying states may influence the observed decay times. PL efficiency in this sample is relatively high at the lowest temperatures and high excitation densities. It drops off rapidly as temperature increases and as excitation density decreases.

In summary, the value of the radiative recombination coefficient B at room temperature is not larger than the corresponding value for bulk GaAs. Results from the MBE MQW show that B is considerably less than $5 \times 10^{-10} \text{ cm}^3 \cdot \text{s}^{-1}$, while results from the MOCVD MQW put a rough upper limit of $7 \times 10^{-11} \text{ cm}^3 \cdot \text{s}^{-1}$ on the value of B . These values are consistent with or lower than the value B for bulk GaAs at low intrinsic carrier densities [23], suggesting that confinement in a 100 Å quantum well does not greatly speed up room temperature free carrier radiative recombination and may in fact slow it down.

Even though B is not greater in quantum wells than it is in bulk GaAs at room temperature, quantum well lasers may still outperform bulk lasers because of the advantageous high carrier densities which can be collected in a single quantum well. In addition, the density of traps per unit volume is lower in the SQW structure than in the MOCVD MQW structure, allowing for easier trap saturation. Hence, the SQW structure exhibits high-luminescence efficiency at room temperature down to reasonably low carrier densities.

These results favor the SQW design over the MQW design for laser active regions. However, the optical confinement factor for single quantum well active layers is poor; the optical mode must fill a volume much larger than the single quantum well gain region. One solution is the graded-index waveguide separate-confinement heterostructure design, which can lower laser thresholds by increasing the optical confinement factor [24]–[25]. A compromise of a few wells appears to be the best solution in the end to achieve high carrier densities in the wells along with good optical confinement [26].

ACKNOWLEDGMENT

We gratefully thank A. C. Gossard for providing the MBE MQW sample and R. C. Miller for providing helium temperature excitation spectra, both of AT&T Bell Laboratories. We also wish to thank M. D. Sturge of Bell Communications and R. C. Miller for their useful comments and discussions.

REFERENCES

- [1] J. E. Fouquet, A. E. Siegman, R. D. Burnham, and T. L. Paoli, "Carrier trapping in room temperature, time-resolved photoluminescence of a GaAs/Al_{0.3}Ga_{0.7}As multiple quantum well structure grown by metalorganic chemical vapor deposition," *Appl. Phys. Lett.*, vol. 46, pp. 374–376, 1985.
- [2] —, "Time-resolved photoluminescence of GaAs/Al_{0.3}Ga_{0.7}As quantum well structures grown by metalorganic chemical vapor deposition," in *Proc. Picosecond Electronics and Optoelectronics*, Incline Village, NV, Mar. 13–15 1985.
- [3] J. E. Fouquet and A. E. Siegman, "Room temperature photoluminescence times in a GaAs/Al_{0.3}Ga_{0.7}As molecular beam epitaxy multiple quantum well structure," *Appl. Phys. Lett.*, vol. 46, pp. 280–282, 1985.
- [4] J. E. Fouquet, "Recombination dynamics in quantum well semiconductor structures," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1986, University Microfilms, Ann Arbor, MI.
- [5] R. C. Miller, A. C. Gossard, G. D. Sanders, Yia-Chung Chang, and J. N. Schulman, "New evidence of extensive valence band mixing in GaAs quantum wells through photoluminescence studies," *Phys. Rev. B*, vol. 32, pp. 8452–8454, 1985.
- [6] P. D. Dapkus, N. Holonyak, Jr., R. D. Burnham, D. L. Keune, J. W. Burd, K. L. Lawley, and R. E. Walline, "Spontaneous and stimulated carrier lifetime (77° K) in a high-purity, surface free GaAs epitaxial layer," *J. Appl. Phys.*, vol. 41, pp. 4194–4199, 1970.
- [7] C. B. Su, R. Olshansky, J. Manning, and W. Powazinik, "Carrier dependence of the radiative coefficient in III-V semiconductor light sources," *Appl. Phys. Lett.*, vol. 44, pp. 732–734, 1984.
- [8] G. Bastard, "Hydrogenic impurity states in a quantum well: A simple model," *Phys. Rev. B*, vol. 24, pp. 4714–4722, 1981.
- [9] B. Lambert, B. Devaud, A. Regreny, and G. Talalaieff, "Impurity photoluminescence in GaAs/Ga_{0.3}Al_{0.7}As multiple quantum wells," *Solid State Commun.*, vol. 43, pp. 443–446, 1982.
- [10] J. I. Pankove, *Optical Processes in Semiconductors*. New York: Dover, 1975.
- [11] J. A. Brum, G. Bastard, and C. Guillemot, "Screened Coulombic impurity bound states in semiconductor quantum wells," *Phys. Rev. B*, vol. 30, pp. 905–908, July 1984.
- [12] N. Holonyak, Jr., R. M. Kolbas, W. D. Laidig, M. Altarelli, R. D. Dupuis, and P. D. Dapkus, "Phonon-sideband MO-CVD quantum-well Al_{0.3}Ga_{0.7}As-GaAs heterostructure laser," *Appl. Phys. Lett.*, vol. 34, pp. 502–505, 1979; N. Holonyak, Jr., R. M. Kolbas, W. D. Laidig, B. A. Vojok, K. Hess, R. D. Dupuis, and P. D. Dapkus, "Phonon-assisted recombination and stimulated emission in quantum well Al_{0.3}Ga_{0.7}As-GaAs heterostructures," *J. Appl. Phys.*, vol. 51, pp. 1328–1336, 1980; N. Holonyak, Jr., R. M. Kolbas, R. D. Dupuis, and P. D. Dapkus, "Quantum-well heterostructure lasers," *IEEE J. Quantum Electron.*, vol. QE-16, pp. 170–186, 1980.
- [13] C. Weisbuch, R. C. Miller, R.ingle, A. C. Gossard and W. Wiegmann, "Intrinsic radiative recombination from quantum states in GaAs-Al_{0.3}Ga_{0.7}As multi-quantum well structures," *Solid State Commun.*, vol. 37, pp. 219–222, 1981.
- [14] C. V. Shank, R. L. Fork, B. I. Greene, C. Weisbuch, and A. C. Gossard, "Picosecond dynamics of highly excited multi-quantum well structures," *Surface Sci.*, vol. 113, pp. 108–111, 1982.
- [15] Z. Y. Xu, V. G. Kreismanis, and C. L. Tang, "Stimulated emission of GaAs-Al_{0.3}Ga_{0.7}As multiple quantum well structures grown by metalorganic chemical vapor deposition," *Appl. Phys. Lett.*, vol. 44, pp. 136–138, 1984.
- [16] S. Tarucha, Y. Horikoshi, and H. Okamoto, "Optical absorption characteristics of GaAs-AlGaAs multi-quantum-well heterostructure waveguides," *J. Appl. Phys.*, vol. 22, pp. L482–L484, 1983.
- [17] R. C. Miller, A. C. Gossard, W. T. Tsang, and O. Munteanu, "Bound excitons in p-doped GaAs quantum wells," *Solid State Commun.*, vol. 43, p. 519–522, 1982.
- [18] J. Hegarty and M. D. Sturge, "Exciton holeburning in GaAs/GaAlAs multiquantum wells," in *Proc. 17th Int. Conf. Physics Semiconductors*, San Francisco, CA, Aug. 6–10, 1984; J. Hegarty, L. Goldner, and M. D. Sturge, "Localized and delocalized two-dimensional excitons in GaAs-AlGaAs multi-quantum well structures," *Phys. Rev. Lett.*, vol. 30, pp. 7346–7348, 1984; M. D. Sturge, J. Hegarty, and L. Goldner, "Localization of the two-dimensional excitons in GaAs-AlGaAs quantum well layers," in *Proc. 17th Int. Conf. Physics Semiconductors*, San Francisco, CA, Aug. 6–10, 1984; M. D. Sturge, private communication.
- [19] P. Dawson, G. Duggan, H. I. Ralph, and K. Woodbridge, "Free

excitons in room-temperature photoluminescence of GaAs-Al_xGa_{1-x}As multiple quantum wells," *Phys. Rev. B*, vol. 28, pp. 7381-7383, 1983.

- [20] D. Bimberg, J. Christen, A. Steckenborn, G. Weimann, and W. Schlapp, "Injection, intersubband relaxation and recombination in GaAs multiple quantum wells," *J. Luminescence*, vol. 30, pp. 562-579, 1985.
- [21] R. C. Miller, W. T. Tsang and O. Munteanu, "Extrinsic layer at Al_xGa_{1-x}As interfaces," *Appl. Phys. Lett.*, vol. 41, pp. 374-376, 1982.
- [22] D. Bimberg, J. Christen, and A. Steckenborn, "Advantages of multiple quantum wells with abrupt interfaces for light emitting devices," in *Two-Dimensional Systems, Heterostructures, and Superlattices*, G. Bauer, F. Kuchar, and H. Heinrich, Eds. Berlin: Springer-Verlag, 1984, pp. 136-146.
- [23] R. J. Nelson and R. G. Sobers, "Minority-carrier lifetime and internal quantum efficiency of surface-free GaAs," *J. Appl. Phys.*, vol. 49, pp. 6103-6108, 1978.
- [24] D. Kasemset, C.-S. Hong, N. B. Patel, and P. D. Dapkus, "Graded barrier single quantum well lasers—theory and experiment," *IEEE J. Quantum Electron.*, vol. QE-19, pp. 1025-1030, 1983.
- [25] W. T. Tsang, "A graded-index waveguide separate-confinement laser with very low threshold and a narrow Gaussian beam," *Appl. Phys. Lett.*, vol. 39, pp. 134-137, 1981; W. T. Tsang, "Extremely low threshold (AlGa)As graded-index waveguide separate-confinement heterostructure grown by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 40, pp. 217-219, 1982; W. T. Tsang and R. L. Hartman, "Electro-optical characteristics of graded-index waveguide separate-confinement heterostructure lasers with proton-delineated stripe," *Appl. Phys. Lett.*, vol. 42, pp. 551-553, 1983.
- [26] J. E. Epler, G. S. Jackson, N. Holonyak, Jr., R. L. Thornton, R. D. Burnham, and T. L. Paoli, "Broadband operation of coupled-stripe multiple quantum well AlGaAs laser diodes," *Appl. Phys. Lett.*, vol. 47, pp. 779-780, 1985.



Julie E. Fouquet (S'84-M'85) was born in California in 1958. She received the Bachelor's degree in physics from Harvard University, Cambridge, MA, in 1980, and the M.S. and Ph.D. degrees in applied physics from Stanford University, Stanford, CA. Her dissertation research topic was recombination dynamics in quantum well semiconductor structures, studied using time-correlated single photon counting.

She joined Hewlett-Packard Laboratories, Palo Alto, CA, in 1985, where she is currently investigating luminescence of a variety of semiconductor materials and devices.

Dr. Fouquet is a member of the IEEE Lasers and Electro-Optics Society and the American Physical Society.



Robert D. Burnham (M'76-SM'83-F'86) was born in Havre de Grace, MD, on March 21, 1944. He received the B.S., M.S., and Ph.D. degrees from the University of Illinois, Urbana, in 1966, 1968, and 1971, respectively.

He held an NDEA Fellowship from 1966 to 1969 and a General Telephone and Electronics Fellowship from 1969 to 1971. He has been a member of the Research Staff at the Xerox Palo Alto Research Center, Palo Alto, CA, since 1971 and is currently a Research Fellow. His work has

been centered on the growth and characterization of III-V compounds for device purposes.

Tunable-Laser-Induced Gratings for the Measurement of Ultrafast Phenomena

RICK TREBINO, CHARLES E. BARKER, AND ANTHONY E. SIEGMAN, FELLOW, IEEE

Abstract—Optically induced gratings form the basis of a class of optical frequency-domain techniques for the measurement of ultrafast phenomena with demonstrated temporal resolution on the order of tens of femtoseconds without the use of ultrashort pulses. We describe these techniques and calculate experimental line shapes. We discuss experimental strategies and some experimental problems that have plagued early experiments, but which have been solved. In addition, we present results obtained for triphenyl methane dyes and the optical Kerr liquid, carbon disulfide, and we discuss the interpretation of these results.

I. INTRODUCTION

WHILE somewhat overshadowed by recent spectacular results using ultrashort pulses [1], [2], tunable-laser-induced-grating techniques employing long-pulse or CW radiation have been obtaining femtosecond temporal resolution for over a decade [3]–[6]. These techniques, which operate in the frequency domain, yield theoretically equivalent information to that obtained by ultrashort-pulse, time-domain methods [7]. But they are not limited by the laser pulse length in their temporal resolution; the measurement of ultrashort time scales requires instead the creation of large optical frequency differences, which is not difficult experimentally. Consequently, the results of femtosecond relaxation studies using tunable-laser-induced-grating, frequency-domain techniques have been no less spectacular than those of ultrashort-pulse, time-domain experiments [8]–[10].

In this review of techniques and results using tunable-laser-induced-gratings for ultrashort-event measurement, we will begin with a description and a short history of such methods and results, discussing the advantages and disadvantages of the various interpretations that have evolved for the methods (Section II). An elementary treatment of the theory behind these techniques follows (Section III). For the measurement of excited-state lifetimes, we employ a simple rate-equation approach [6], [7], which is valid when the material lifetimes are much longer than the dephasing time (an appropriate assumption for most of the experiments performed to date). An advantage of this simple approach is that it allows saturation effects to be treatable in second order [7] rather than in fifth order as in density-matrix calculations [11]. We also discuss the

problem of induced-grating formation using orthogonally polarized excitation beams [12]. Our treatment of the theory concludes with a simple derivation of the tunable-laser-induced-grating diffraction-efficiency line shape as the Fourier transform of the material impulse response [7]. We continue with a discussion of experimental strategy (Section IV). The choice of beam geometry represents an important problem here, and we discuss this problem in detail. We pay particular attention to the issues of maintaining phase matching during long wavelength scans [13] and whether a two- or three-laser geometry is preferable. We also discuss the choice of experimental parameters, the effects of Doppler broadening, and avoidance of asymmetrical line shapes, and the problem of curve fitting the resulting experimental line shapes. We proceed to describe the experimental apparatus that we have used to perform tunable-laser-induced-grating experiments for measuring femtosecond lifetimes (Section V). Results that we have obtained in experiments on nonfluorescent organic dyes comprise Section VI, and the problematic issue of thermal gratings follows in the next section, which also contains a discussion of thermal-grating suppression techniques. Finally, we discuss some recent results on the optical-Kerr material, carbon disulfide (Section VIII), whose femtosecond orientational-relaxation component has been the subject of intense study; our data for this material are intriguing and far from well understood.

II. THE TECHNIQUE AND ITS EVOLUTION

The use of frequency-domain (tunable-laser-induced-grating) techniques for the measurement of ultrafast phenomena evolved almost simultaneously within three separate spectroscopic disciplines: induced-grating spectroscopy [6], saturation spectroscopy [5], [9], and four-wave mixing [3]. Consequently, numerous names for the method have resulted, ranging from the simple "polarization spectroscopy" [5] to "resonant Rayleigh-type optical mixing" [8] to "nearly degenerate four-wave mixing" [14]. Because the induced-grating interpretation of this class of techniques is so intuitive and productive, we have adopted the name the "tunable-laser-induced grating" technique, which we shall employ throughout this work. In any case, the techniques are all members of the same class of optical frequency-domain techniques, and they differ only in details, usually experimental.

Researchers in the field of saturation spectroscopy (see Fig. 1) were actually the first to realize the potential of

Manuscript received November 25, 1985; revised February 17, 1986. This work was supported by the U.S. Air Force Office of Scientific Research.

The authors are with the Department of Applied Physics, Stanford University, Stanford, CA 94305.

IEEE Log Number 8608669.



Fig. 1. The beam geometry of saturation spectroscopy and of early tunable-laser-induced-grating experiments. The two beams labeled ω_2 in this figure represent one excitation beam and the probe beam, but are incorporated into a single laser beam originating from the same laser. In this figure and in Figs. 2-5, solid lines indicate the input beams (i.e., the excitation and the probe beams) and dashed lines indicate the diffracted "signal" beams. Polarizations are indicated when relevant.

these frequency-domain techniques for the measurement of relaxation times. In 1964, Lamb [15], in his classic analysis of one of the most famous saturation phenomena, the "Lamb dip," was naturally led to consider the interaction of slightly different frequency beams (i.e., different laser modes) in the laser gain medium. The Lamb dip itself does not require coherent beams, but Baklanov and Chebotaev [16], [17] (1971) extended Lamb's analysis by including coherent-coupling effects and demonstrated the existence of an additional narrower absorption dip whose width depended only on longitudinal material relaxation times. By correctly interpreting this dip as due to the intensity modulation at the frequency difference between the two beams, and in noting that lifetimes could be measured by studying the dip, Baklanov and Chebotaev became the first researchers to announce what was essentially a tunable-laser-induced-grating method. Interestingly, they recommended the technique for the measurement of *long* lifetimes.

Sargent and co-workers [18]-[20] (1976), on the other hand, suggested the method for the measurement of very *short* lifetimes. But more importantly, they realized that the extra dip derived from Bragg scattering from induced gratings, and they proceeded to study the effects of these gratings inside laser gain media. Keilmann and co-workers [4], [9], [21]-[24] (1977) applied the "tunable-laser-induced-grating dip" technique to the problem of measuring hot-hole relaxation times in p-type germanium with CO₂ laser radiation, and observed hot-hole-optical-phonon scattering times as short as 0.11 ps. Herman *et al.* [23] (1978) and Wiggins and co-workers [26]-[28] (1978), in studying surfaces, considered the similar problem of the change in *reflectivity* of a weak beam in the presence of a strong beam of slightly different frequency, recognizing the grating contribution of the variation in reflectivity. The latter group performed experiments on germanium surfaces looking at thermal relaxations.

Because they studied the change in material absorption or reflection of a weak beam (in the presence of a strong beam), the researchers mentioned above were required to measure small changes in a weak beam's intensity—changes that were invariably already present in the form of laser noise. As a result, early experimenters necessarily confined themselves to highly saturable systems, i.e., semiconductors, which yielded changes in absorption due to this effect of greater than 1 percent [9], [28]. By basing their variation of the method on polarization spectroscopy, a nearly zero background variation of saturation spectroscopy (see Fig. 2), Song and co-workers [5], [29],



Fig. 2. The beam geometry of polarization spectroscopy.

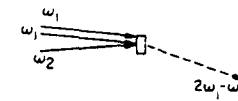


Fig. 3. The beam geometry of "resonant Rayleigh-type optical mixing" employed by Yajima and co-workers.

[30] improved the sensitivity of the technique by several orders of magnitude. They simply rotated the polarization of their signal beam (the radiation producing the "dip") by rotating the polarization of the strong beam so that the background intensity of the weak beam could be filtered out with a polarizer. As a result, they were able to study systems such as organic dyes, yielding much weaker signals.

In all of the above saturation-spectroscopic methods, the signal beam possessed the same frequency and k vector as the weak beam, thus appearing as a change in absorption in the method of Baklanov and Chebotaev [16], [17] and later researchers [18]-[24], as a change in reflection in the method of Herman *et al.* [25] and Wiggins and co-workers [26]-[28], and as a change in polarization in the method of Song, Lee, and Levenson [5], [29], [30]. In the language of four-wave mixing, the frequency and k -vector relations would be written

$$\omega_{\text{sig}} = (\omega_1 - \omega_2) + \omega_2 \quad (1)$$

$$\vec{k}_{\text{sig}} = (\vec{k}_1 - \vec{k}_2) + \vec{k}_2 \quad (2)$$

where ω_{sig} , ω_1 , and ω_2 are the frequencies of the signal, weak, and strong waves, respectively, with the k -vectors similarly defined. The parenthetical quantities in (1) and (2) are the induced-grating frequency and k -vector, respectively. It is clear from these equations that phase matching (or equivalently, satisfaction of the Bragg condition) occurs automatically—with $\omega_{\text{sig}} = \omega_1$ and $k_{\text{sig}} = k_1$ —and as a result, saturation-spectroscopic techniques obscure the four-wave-mixing nature of the process.

Conscious of the line narrowing that often occurs in nonlinear optical spectra, researchers in the field of four-wave mixing soon realized that, depending upon which resonant contribution to the third-order susceptibility is measured, narrowing all the way to a natural line width $1/\tau$ or merely to a dephasing width $1/T_2$ would occur. Thus, by observing the dispersion of a material's $\chi^{(3)}$, interesting information regarding the various relaxation rates could be obtained.

Yajima and co-workers [3], [8], [11] (1975) were the first to take advantage of these four-wave-mixing ideas to measure relaxation times. Looking at processes (see Fig. 3) of the form

$$\omega_{\text{sig}} = 2\omega_1 - \omega_2 \quad (3)$$

with $\omega_1 - \omega_2 \approx 0$, they measured various longitudinal

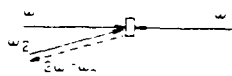


Fig. 4. The beam geometry of phase conjugation by four-wave mixing.



Fig. 5. The beam geometry of tunable-laser-induced-grating experiments involving three separate input beams.

and transverse lifetimes of organic dyes in solids and in liquid solutions, obtaining values as short as 0.02 ps. Raj and co-workers [31], [32] (1980) studied atomic hydrogen and molecular iodine, also using $2\omega_1 - \omega_2$ processes, but employing the phase-conjugate geometry (see Fig. 4) and a heterodyne detection scheme in addition. Other researchers employing $2\omega_1 - \omega_2$ wave-mixing techniques include Yuen and Wolf [33] (1982) and also Kenney-Wallace, Wallace, and Golombok [34]–[38] (1982–1985), who first used four-wave mixing in the phase-conjugate geometry to study the influence of changing microscopic molecular interactions on the ultrafast orientational relaxation times of probe molecules in liquids.

Hochstrasser and co-workers [39] (1978) extended these techniques to coherent anti-Stokes Raman spectroscopy, in which $\omega_{\text{sig}} = 2\omega_1 - \omega_2$ again, but $\omega_1 - \omega_2 \approx \omega_L$, where ω_L is the frequency of a Raman transition. Such a method avoids the problem of thermal gratings (see Section VII), first observed and elucidated by Andrews and Hochstrasser [40], [41], but always results in a line shape with some transverse-relaxation broadening [42]. Under some conditions, this additional broadening is less than the absorption linewidth [43], but in general, only lower limits on the actual lifetime are available with this method.

In the saturation-spectroscopic interpretation, the presence of the wavelength-tunable beam alters the material absorption of the fixed-frequency beam, while in the four-wave-mixing view, one or more terms of the material third-order susceptibility act to produce simultaneous sum- and difference-frequency generation. The induced-grating picture, we feel, yields a more intuitive explanation of the physics of the process. In the induced-grating interpretation, two excitation beams have a tunable frequency difference $\Delta\omega = \omega_1 - \omega_2$ and interfere in a sample material to produce a moving interference pattern with frequency $\Delta\omega$. This pattern creates a matching excited-state-grating pattern in the sample, which diffracts the probe beam. No delay is introduced as in time-domain techniques; the frequency-domain nature of the technique requires CW or near-CW beams, and hence, the grating is created and probed continuously and simultaneously. The diffracted intensity varies according to the frequency difference $\Delta\omega$ and the material relaxation rate $1/\tau$. When $|\Delta\omega|$ is much greater than $1/\tau$, the intensity fringes sweep through too quickly for the material to follow. The material grating washes out and diffracts very little probe light. On the other hand, when $|\Delta\omega|$ is much less than $1/\tau$, the intensity fringes appear nearly stationary to the material, i.e., each molecule sees the intensity oscillating at a rate slower than its response. The material can follow the intensity fringes easily, resulting in maximal diffraction efficiency for the induced grating. We thus expect the cutoff between the high- and low-diffraction-efficiency re-

gimes to occur at $|\Delta\omega| \approx 1/\tau$. A "tunable-laser-induced-grating" experiment, then, involves varying $\Delta\omega$, measuring the induced-grating diffraction efficiency as a function of $\Delta\omega$, and determining the material lifetime by finding the half-width half-maximum of the diffraction-efficiency line shape: $\tau \approx 1/\Delta\omega_{\text{HWHM}}$. In Section III, we will show that, for a simple exponential decay, this relation is exact. More generally, the diffraction-efficiency line shape will be the Fourier transform of the material impulse-response function. This result will be seen to follow quite simply in the induced-grating interpretation.

The development of the technique within the field of induced-grating spectroscopy occurred several times. As mentioned earlier, Sargent and co-workers [18]–[20] and Keilmann and co-workers [4], [9], [21]–[24] both realized that induced gratings played an important role in their techniques. In addition, Siegman, who had been working with a time-domain, ultrashort-pulse induced-grating technique for measuring excited-state lifetimes [44], realized that *moving* gratings could also yield excited-state-lifetime information. He suggested a method in which two separate lasers form a dynamic grating, and a third laser at an additional wavelength probes the grating [6].

While the induced-grating picture does not require the use of a third wavelength for the probe, it is useful to consider techniques involving three input frequencies (see Fig. 5). In a four-wave-mixing interpretation, such a technique involves processes of the form

$$\omega_{\text{sig}} = \omega_1 - \omega_2 + \omega_3. \quad (4)$$

The versatility of such an arrangement will be seen to be quite advantageous.

It is worthwhile to keep in mind all of these interpretations of the technique to attain maximal understanding of experimental data. In particular, the four-wave-mixing interpretation provides a good formalism for describing measurable quantities, while the induced-grating interpretation provides an excellent intuitive picture of the process. These two interpretations, in naturally allowing general processes of the form $\omega_{\text{sig}} = \omega_1 - \omega_2 + \omega_3$, are especially useful.

III. THEORY

A. A Three-Level Rate-Equation Description of the Tunable-Laser-Induced-Grating Technique for the Measurement of Excited-State Lifetimes

In this section, we derive the basic theoretical expressions required for tunable-laser-induced-grating studies of excited states in simple systems. The results of this section provide an intuitive appreciation for the method that will be useful for understanding more complex systems. We treat a three-level system excited and probed by mono-

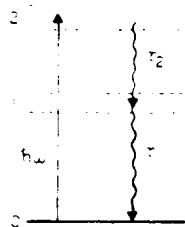


Fig. 6. Three-level system with relaxation times.

chromatic plane waves. A rate-equation analysis involving an iterative solution procedure provides the simple mathematical machinery necessary to understand the method. Here, we perform only the first step of this iterative procedure, the first two thirds of which is equivalent to the results of Siegman [6]; reference [7] performs the second step of this procedure, which yields the saturation behavior of the method in a simple manner.

Consider the three-level system of Fig. 6, with level two decaying at a rate $1/\tau_2$ primarily to level one and with level one decaying at a rate $1/\tau_1$ to the ground level zero. We assume that various (homogeneous and/or inhomogeneous) broadening mechanisms [45] broaden level two so that both excitation frequencies ω_1 and $\omega_2 = \omega_1 - \Delta\omega$ resonantly connect the ground state to this excited level. The excitation fields will be assumed to have the form

$$E_1(r, t) = \text{Re } E_1 \exp [i(\omega_1 t - \vec{k}_1 \cdot \vec{r})] \hat{e}_{\text{ex}} \quad (5)$$

$$E_2(r, t) = \text{Re } E_2 \exp [i(\omega_2 t - \vec{k}_2 \cdot \vec{r})] \hat{e}_{\text{ex}}. \quad (6)$$

These fields simultaneously illuminate the above system producing the time-averaged intensity

$$I(r, t) = I_1 + I_2 + \text{Re} \{ (c/4\pi)(\epsilon/\mu)^{1/2} E_1 E_2^* \cdot \exp [i(\Delta\omega t - \vec{k}_{gr} \cdot \vec{r})] \} \quad (7)$$

where $I_i = (c/8\pi)(\epsilon/\mu)^{1/2} |E_i|^2$ is the intensity [46] of the i th excitation beam, $\vec{k}_{gr} = \vec{k}_1 - \vec{k}_2$ is the induced-grating k vector, and rapidly oscillating terms with frequencies $2\omega_1$, $2\omega_2$, and $\omega_1 + \omega_2$ have been neglected.

In the absence of coherence effects, rate equations suffice to describe the system

$$\frac{d}{dt} N_2(r, t) = \frac{\sigma_{\text{ex}}}{\hbar\omega_1} I(r, t) N_0(r, t) - \frac{1}{\tau_2} N_2(r, t) \quad (8)$$

$$\frac{d}{dt} N_1(r, t) = \frac{1}{\tau_2} N_2(r, t) - \frac{1}{\tau_1} N_1(r, t) \quad (9)$$

$$\frac{d}{dt} N_0(r, t) = \frac{1}{\tau_1} N_1(r, t) - \frac{\sigma_{\text{ex}}}{\hbar\omega_1} I(r, t) N_0(r, t) \quad (10)$$

$$N = N_0(r, t) + N_1(r, t) + N_2(r, t) \quad (11)$$

where $N_0(r, t)$, $N_1(r, t)$, and $N_2(r, t)$ are the population densities of the various levels, σ_{ex} is the absorption cross section at the excitation wavelength, and N is a constant representing the total molecular density of the material under study. We neglect the effects of stimulated emission

on the level populations, and we also neglect the effects of absorption and depletion of the excitation-beam intensities.

If we assume that only a small fraction of the ground-state molecules are excited in the process, then the following one-step iterative solution to (8)–(11) will be appropriate: first approximate the ground-state population density $N_0(r, t)$ by the total population density N (a constant), and solve (8) and (9) for $N_2(r, t)$ and $N_1(r, t)$; then substitute these newly derived values into (11) to obtain $N_0(r, t)$. Siegman [6] performed the first two thirds of this procedure, obtaining $N_1(r, t)$ and $N_2(r, t)$. The iteration may continue, i.e., the derived value of $N_0(r, t)$ could then be used in (8) and (9) again to improve the estimates of $N_2(r, t)$ and $N_1(r, t)$, etc., but the second step will prove unnecessary unless saturation effects are important (see [7]).

Performing this one-step procedure, we easily find

$$N_2(r, t) = \frac{N\sigma_{\text{ex}}\tau_2}{\hbar\omega_1} \left[I_1 + I_2 + \text{Re} (c/4\pi)(\epsilon/\mu)^{1/2} E_1 E_2^* \cdot \frac{\exp [i(\Delta\omega t - \vec{k}_{gr} \cdot \vec{r})]}{1 + i\Delta\omega\tau_2} \right] \quad (12)$$

$$N_1(r, t) = \frac{N\sigma_{\text{ex}}\tau_1}{\hbar\omega_1} \left[I_1 + I_2 + \text{Re} (c/4\pi)(\epsilon/\mu)^{1/2} E_1 E_2^* \cdot \frac{\exp [i(\Delta\omega t - \vec{k}_{gr} \cdot \vec{r})]}{(1 + i\Delta\omega\tau_1)(1 + i\Delta\omega\tau_2)} \right] \quad (13)$$

and

$$N_0(r, t) = N - \frac{N\sigma_{\text{ex}}\tau_0}{\hbar\omega_1} \left[I_1 + I_2 + \text{Re} (c/4\pi)(\epsilon/\mu)^{1/2} E_1 E_2^* \cdot \frac{(1 + i\Delta\omega\tau_{12}) \exp [i(\Delta\omega t - \vec{k}_{gr} \cdot \vec{r})]}{(1 + i\Delta\omega\tau_1)(1 + i\Delta\omega\tau_2)} \right] \quad (14)$$

where, in (14), $\tau_{12}^{-1} = \tau_1^{-1} + \tau_2^{-1}$, and the "ground-state-recovery time" τ_0 is given by $\tau_0 = \tau_1 + \tau_2$. Thus, all levels experience traveling-wave modulations that are maximal when $\Delta\omega = 0$ and minimal when $|\Delta\omega| \rightarrow \infty$.

Any of these population-density modulations will diffract a simultaneously incident probe beam chosen resonant among level zero, one, or two and another level of the molecule. In the low-efficiency limit, the diffraction efficiency due to an absorption grating in the i th level η_i can be shown, using a standard slowly varying envelope approximation to Maxwell's equations for plane waves in macroscopic media, assuming a phase-matched interaction, to be

$$\eta_i = |\Delta N_i \sigma_{\text{pr}} L/2|^2 \quad (15)$$

where ΔN_i is the amplitude of the oscillatory component of the population-density modulation of the probed level, σ_{pr} is the absorption cross section of the probed level at the probe wavelength, and L is the sample thickness. Substitution of the oscillatory components of the population

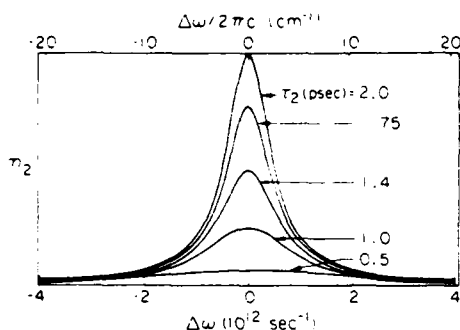


Fig. 7. Theoretical tunable-laser-induced-grating lineshape obtained by probing the induced grating in the second excited level of the system shown in Fig. 6.

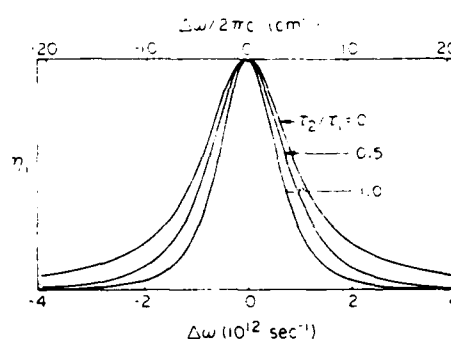


Fig. 8. Theoretical tunable-laser-induced-grating line shape obtained by probing the induced grating in the first excited level of the system shown in Fig. 6.

densities from (12), (13), and (14) into (15) yields the diffraction efficiency of a grating in the i th level (indicated by the subscript):

$$\eta_2(\Delta\omega) = [N\sigma_{\text{ex}}\sigma_{\text{pr}}L/\hbar\omega_1]^2 I_1 I_2 \tau_2^2 / (1 + i\Delta\omega\tau_2)^2 \quad (16)$$

$$\eta_1(\Delta\omega) = [N\sigma_{\text{ex}}\sigma_{\text{pr}}L/\hbar\omega_1]^2 I_1 I_2 \tau_1^2 / [(1 + i\Delta\omega\tau_1)(1 + i\Delta\omega\tau_2)]^2 \quad (17)$$

$$\eta_0(\Delta\omega) = [N\sigma_{\text{ex}}\sigma_{\text{pr}}L/\hbar\omega_1]^2 I_1 I_2 \tau_1^2 (1 + i\Delta\omega\tau_{12}) / [(1 + i\Delta\omega\tau_1)(1 + i\Delta\omega\tau_2)]^2 \quad (18)$$

where the functional dependence of the diffraction efficiency on $\Delta\omega$ has been emphasized by writing $\eta_i(\Delta\omega)$.

From (16), we see that probing level two results in a line shape (diffraction efficiency versus $\Delta\omega$ curve) that is Lorentzian (see Fig. 7) with a half-width half-maximum ($\Delta\omega_{\text{HWHM}}$) of $1/\tau_2$, i.e., $\Delta\omega_{\text{HWHM}} \cdot \tau_2 = 1$. Note that $\eta_2(\Delta\omega)$ is independent of τ_1 : when level two is probed, ground- and first-excited-state molecules are assumed to have negligible absorption cross sections at the probe frequency, and hence appear identical to the probe beam; as a result, the decay rate between these two levels cannot appear in the expression for the second-excited-state diffraction efficiency. In the limit $\tau_1 \rightarrow \infty$, however, a bottleneck in the first excited state will occur, resulting in ground-state depletion. A second iteration in the rate-equation solution procedure would then be required, resulting in a somewhat reduced level-two population density, and hence a lessened diffraction efficiency. This effect is essentially just transition saturation, and from (14), we see that it occurs when $I_1 + I_2$ approaches $\hbar\omega_1/\sigma_{\text{ex}}\tau_0$, which is not surprising.

Another characteristic of the level-two line shape is that the peak diffraction efficiency $\eta_2(0)$ is proportional to τ_2^2 : the population density of level two is proportional to τ_2 since the longer the lifetime of this excited state, the more molecules will remain in this level; the diffraction efficiency is then proportional to the square of the population density. This dependence illustrates a limitation of the technique: the diffraction from a sample with a very short lifetime will be very weak.

A more complicated curve—the product of two Lorentzians—is obtained if level one is probed (see Fig. 8).

Generally speaking, the level-one diffraction efficiency appears Lorentzian with a half-width half-maximum equal to the inverse of the longer lifetime, but in the tails beyond the inverse of the shorter lifetime, the asymptotes approach zero as $\Delta\omega^{-4}$ instead of $\Delta\omega^{-2}$. This double-Lorentzian behavior results from two effects: the washing out of the first-excited-state grating when $|\Delta\omega| > 1/\tau_1$, and also the washing out of the second-excited-state grating because the first-excited-state population is provided solely by the decaying second-excited-state population whose distribution will cease to be strongly sinusoidal when $|\Delta\omega| > 1/\tau_2$.

Probing the ground state presents the most complicated line shape (18). If, however, one lifetime is much less than the other, the short lifetime cancels out of the line shape expression, and only the longer lifetime is measured. This lifetime can then be interpreted as the "ground-state-recovery time" as it is the rate-limiting step in the decay from the second excited state to ground. Intermediate situations, in which the lifetimes are comparable, pose difficulties in interpretation unless other information, perhaps from probing other levels, is available.

Since τ_2 appears in both level-one and level-two diffraction efficiencies, either level may be probed to determine its value. If $\tau_2 \ll \tau_1$, a comparison of (16) and (17) indicates that the diffraction efficiencies that must be detected to observe effects due to τ_2 are exactly equal for the two levels. On the other hand, if $\tau_2 \gg \tau_1$, the required detectable diffraction efficiency obtained by probing level two exceeds that of level one by $(\tau_2/\tau_1)^2$, so in this case, level two should be probed. Of course, the simpler line shape obtained by probing level two is an additional advantage.

The above results follow from the assumptions of the appropriateness of a three-level system and the rate equations in the weak-excitation-beam limit. Authors have treated more complex molecular systems: Trebino [7] and Oudar and Shen [47] treat general N -level systems. The former analysis includes saturation effects, while the latter employs a density matrix approach including dephasing effects. Other authors have also included dephasing effects in tunable-laser-induced-grating calculations [3]–[5], [11], [14].

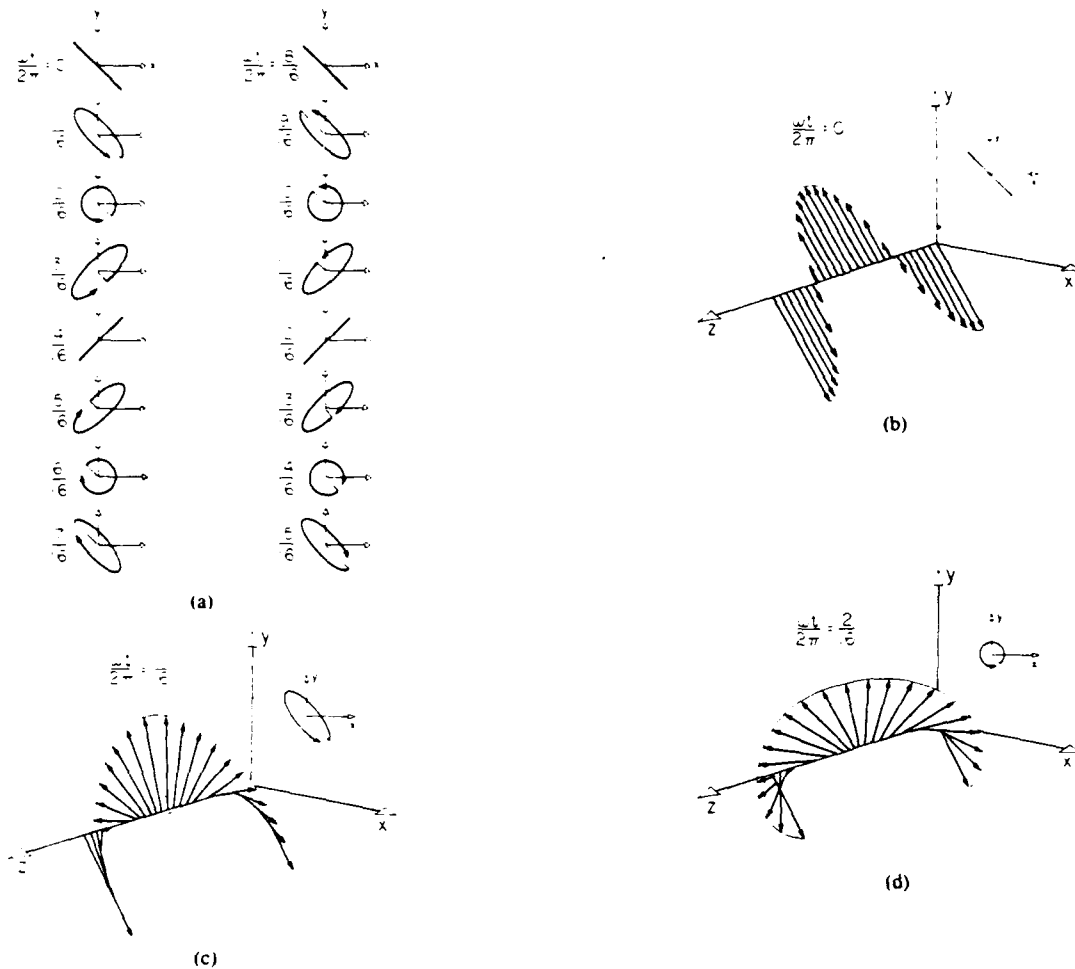


Fig. 9. Instantaneous total electric-field vector due to the superposition of two equal-amplitude, equal-frequency, orthogonally polarized plane waves counterpropagating along the z axis at various times during an optical cycle. Fig. 9(a) depicts the cross-sectional shapes of the spatial envelope of the electric field vectors at various times during an optical cycle. Note that the envelope oscillates from planar to helical and back to planar several times during a cycle. The arrows indicate the handedness of the helical envelopes. Fig. 9(b) shows one spatial period of the electric field at $\omega t/2\pi = 0$. The envelope is a cosine lying in a plane tilted 45° to the x and y axes as shown in the endview inset. Fig. 9(c) shows one spatial period of the electric field at $\omega t/2\pi = 1/6$. At this instant in time, the envelope is a right-handed helix of elliptical cross section. Fig. 9(d) shows one spatial period of the electric field at $\omega t/2\pi = 2/3$. The envelope now is a right-handed helix of circular cross section. As time progresses through one optical cycle, this pattern shifts along the z axis, and the handedness of the helical envelopes reverses four times. The net result of this complicated electric field is a pattern in which the intensity is uniform, but a grating is formed in the electric field direction: there are regions where the electric field direction remains constant and the amplitude varies, and there are regions where the amplitude remains constant, but the field direction rotates through 2π rad during an optical cycle.

B. Orthogonally Polarized Excitation Beams

The derivation in the preceding section required excitation beams with the same polarizations to interfere to produce a spatially sinusoidal component in the intensity. Intensity-dependent effects clearly require such a component to produce a material grating to diffract light. The use of *orthogonally polarized* excitation beams, on the other hand, results in a uniform intensity distribution and, hence, produces no intensity-dependent gratings. In fact, no *scalar* gratings of any kind can be produced with orthogonally polarized excitation beams, which can produce only uniform scalar quantities. *Vector* and *tensor* quantities can, however, experience spatially sinusoidal modulations and, as a result, can cause the diffraction of light [12]. Thus, while the material temperature distri-

bution (a scalar quantity) must necessarily remain uniform, electric- (or magnetic-) field-dependent vector or tensor quantities, such as the induced dipole moment or polarizability, can experience spatially sinusoidal variations and can be studied. Fig. 9 shows the total instantaneous electric field vector produced by orthogonally polarized excitation beams for several times during an optical cycle.

For describing induced-grating experiments using excitation beams of parallel or orthogonal polarizations, it is useful to employ the language of four-wave mixing, that is, $\chi_{ijkl}^{(3)}(-\omega_0; \omega_1, \omega_2, \omega_3)$ where ω_1, ω_2 , and ω_3 are the input frequencies, ω_0 is the output frequency, and i, j, k , and l represent the polarizations of the respective beams. Thus, if ω_1 and ω_2 are the excitation frequencies, $\chi_{xxyy}^{(3)}$ represents excitation with two x -polarized beams, probing

TABLE I
THE DISTINCT, NONZERO THIRD-ORDER-SUSCEPTIBILITY TENSOR
ELEMENTS FOR AN ISOTROPIC MATERIAL AND THEIR
INTERPRETATIONS IN THE INDUCED-GRATING PICTURE*

Tensor Element	Signal Polarization	Excitation Polarizations	Probe Polarization
$\chi_{xxxx}^{(3)}$	\hat{x}	Parallel (\hat{x}, \hat{x})	\hat{x}
$\chi_{xyxx}^{(3)}$	\hat{y}	Parallel (\hat{x}, \hat{x})	\hat{y}
$\chi_{xxyx}^{(3)}$	\hat{y}	Orthogonal (\hat{x}, \hat{y})	\hat{x}
$\chi_{xyyx}^{(3)}$	\hat{x}	Orthogonal (\hat{x}, \hat{y})	\hat{y}

*All other elements are either equal to one of the above elements, or are zero. The subscript order is signal (ω_s), variable-frequency excitation (ω_1), fixed-frequency excitation (ω_2), and probe (ω_3).

with a y-polarized beam, and detection of y-polarized signal radiation. We will continue to take ω_1 and ω_2 to be the frequencies of the excitation beams, so that $\chi_{xxxx}^{(3)}$ and $\chi_{xyxx}^{(3)}$ represent experiments with parallel excitation beam polarizations, while $\chi_{xxyx}^{(3)}$ and $\chi_{xyyx}^{(3)}$ represent experiments with orthogonally polarized excitation beams (see Table I). For isotropic materials, these four elements are the only distinct, nonzero third-order-susceptibility-tensor elements; all others are either zero or are equal to one of the above four [48].

The use of orthogonally polarized excitation beams is important in tunable-laser-induced-grating experiments for several reasons. First, as we will show in Section VII, it will be necessary to suppress temperature gratings, whose diffracted radiation often obscures the radiation diffracted from a more interesting grating, in absorbing samples. In addition, even in nonabsorbing samples, it will be of interest to probe such gratings, which contain additional physical information about the material (see Section VIII).

C. The Diffracted Line Shape as the Fourier Transform of the Material Response

We have seen that the linewidth of the tunable-laser-induced-grating diffraction-efficiency curve indicates the time scale of the rate-limiting process of the phenomenon under study. This simple and useful result is, however, only one aspect of the information obtainable with the technique: in addition, the line shape of the diffraction-efficiency curve indicates the functional form of the decay process. The simple three-level system of the preceding section admitted various line shapes depending on the level or levels probed; the purpose of this section is to develop a simple general interpretation of the tunable-laser-induced-grating line shape.

Suppose that, instead of the CW excitation that we have been considering, an ultrashort, δ -function-like pulse of radiation or other energy perturbs a system. Various changes in the system will occur as a result. Suppose that one of these changes is unobtrusively probed, resulting in the measurement of a system response function $h(t)$ known in linear-system theory as the "impulse response" of the system [49]. The transient-grating technique [50] and

other short-pulse excite-probe techniques [1], [2], measure this function (or its magnitude squared) directly. If, however, the excitation takes a different form, say $I(t)$, then for linear, time-invariant systems, the observed system response $O(t)$ will be given by the convolution of $h(t)$ and $I(t)$:

$$O(t) = \int_{-\infty}^{\infty} h(t') I(t - t') dt'. \quad (19)$$

The excitation in the tunable-laser-induced-grating techniques is

$$I(t) \propto \exp(i\Delta\omega t). \quad (20)$$

Substitution into (19) gives

$$O(t) \propto \exp(i\Delta\omega t) \Gamma(\Delta\omega) \quad (21)$$

where

$$\Gamma(\Delta\omega) = \int_{-\infty}^{\infty} h(t') \exp(-i\Delta\omega t') dt'. \quad (22)$$

In other words, the system response in the tunable-laser-induced-grating technique (i.e., the diffracted field) will be proportional to the Fourier transform of the system-impulse-response function $h(t)$ evaluated at $\Delta\omega$. The complex exponential in (21) merely indicates the Doppler shift suffered by the signal beam with respect to the probe beam because diffraction occurs off moving fringes. The complete time dependence of the diffracted field is obtained simply by multiplying by the probe field, which is proportional to $\exp(i\omega_3 t)$ —integration of Maxwell's equations for this process in the low-diffraction-efficiency limit merely yields proportionality constants [7].

The above derivation can be extended to the case of four-wave-mixing interactions if we can write the four-wave-mixing response function $R(t_0, t_1, t_2, t_3)$ in the form

$$R(t_0, t_1, t_2, t_3) = \delta(t_0 - t_3) h(t_3 - t_2) \delta(t_2 - t_1) \quad (23)$$

where t_0 is the signal beam time, t_1 and t_2 are the excitation beam times, and t_3 is the probe time. The above expression describes any phenomenon that requires the two excitation beams incident simultaneously (i.e., to induce the grating) and that produces a signal beam simultaneously with the probe beam's incidence time. The function $h(t)$ is the impulse response of the material that was considered in the preceding derivation. The frequency response of the material $\chi^{(3)}(\omega_1, -\omega_2, \omega_3)$ will be the three-dimensional Fourier transform of the response function [48]

$$\begin{aligned} \chi^{(3)}(\omega_1, -\omega_2, \omega_3) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t_0, t_1, t_2, t_3) \exp \\ & -i(\omega_1 t_1 - \omega_2 t_2 + \omega_3 t_3) dt_1 dt_2 dt_3 \end{aligned} \quad (24)$$

which is easily evaluated for the response function above to yield

$$\chi^{(3)}(\omega_1, -\omega_2, \omega_3) = \exp i(\omega_1 - \omega_2 + \omega_3)t_0 \Gamma(\Delta\omega) \quad (25)$$

as expected. Equation (25) can be Fourier transformed with respect to t_0 to yield the frequency spectrum of the diffracted radiation. We find

$$\chi^{(3)}(-\omega_0, \omega_1, -\omega_2, \omega_3) = \delta(\omega_0 - \omega_1 + \omega_2 - \omega_3) \Gamma(\Delta\omega) \quad (26)$$

which includes an energy-conserving delta function and the line shape factor $\Gamma(\Delta\omega)$.

IV. EXPERIMENTAL STRATEGY AND CURVE-FITTING TECHNIQUES

Numerous experimental strategies have emerged for performing tunable-laser-induced-grating experiments over the years, ranging from the simple observation of power transfer between two beams in an absorbing medium [4] to the measurement of the four-wave-mixing signal produced by the interaction of beams from three separate lasers in a complex four-beam geometry [10]. In this section, we discuss experimental strategies for such experiments, considering such issues as the optimal lasers and pulse lengths, the best beam geometries, whether to use two or three independent lasers for the experiment, the choice of experimental parameters, and some curve-fitting issues.

We begin with the choice of the appropriate laser. Nearly all experiments so far have employed *Q*-switched-laser-pumped pulsed dye lasers because such lasers combine the advantages of wavelength availability, tunability, high peak power, and nearly optimal pulse length. In addition, their repetition rate is only limited by that of the pump laser. Such lasers provide useful tuning ranges of as much as $\approx 500 \text{ cm}^{-1}$ and line widths as small as $\approx 0.005 \text{ cm}^{-1}$, corresponding to the potential measurement of lifetimes between $\approx 10 \text{ fs}$ and $\approx 1 \text{ ns}$. While pulsed dye lasers probably are the optimal sources available for frequency-domain ultrafast-event measurement, there are, however, some unfortunate characteristics of these lasers that contribute to less than perfect experimental performance. We mention these characteristics here to illustrate some of the experimental problems of the technique, but also to encourage experimenters to try new laser sources as they become available.

Shorter pulse lengths than the few nanoseconds available from these lasers may be advantageous. While the technique requires that the laser pulse length be much longer than the material lifetime(s), material damage thresholds increase with shorter pulse length and the signal strength scales with the input laser intensity cubed. Consequently, the measurement of 100 fs lifetimes might best be done with $\approx 10 \text{ ps}$ pulses.

In addition, transversely pumped pulsed dye lasers generally yield multispatial-mode output. And worse, temporal-Fourier-transform-limited pulses are not generally available from these lasers simultaneously with large tuning range. Poor spatial mode quality and the severe temporal intensity fluctuations within each pulse both act to cause large shot-to-shot jitter in the diffracted-pulse en-

ergy. As a result, we have observed signal-energy shot-to-shot jitter as high as 100 percent, despite input-laser-energy shot-to-shot jitter as low as a few percent. Spatial filtering decreases this noise somewhat, but it is difficult to rid these lasers of their temporal intensity fluctuations while simultaneously maintaining continuous tuning over $\approx 100 \text{ cm}^{-1}$. Such noise underscores the importance of using a background-free geometry for performing these experiments: large shot-to-shot jitter in a background (present in most dynamical-grating geometries) can obscure the often weak line containing the lifetime information.

Numerous geometries have been employed to perform ultrafast-relaxation measurements using dynamical gratings. Most have used two lasers: one at a fixed frequency and the other tunable within $\approx 100 \text{ cm}^{-1}$ about the former frequency. In doing so, these geometries combine one excitation beam and the probe beam into a single laser beam, thus forcing excitation and probing to occur at the same frequency. The alternative is to employ three separate lasers allowing probing to occur at an independent wavelength. Clearly, two-laser geometries allow experimental simplicity, while three-laser arrangements yield greater experimental versatility.

There are, however, other advantages associated with the use of three lasers. First, three-laser geometries achieve background-free operation, while two-laser arrangements generally do not. And second, the freedom of choice of the probe wavelength allows much simpler interpretation of experimental results, often reducing the appropriate theoretical expression to a single term, while two-laser arrangements require several terms and, as a result, incur a much more complex curve-fitting problem.

Consider first the problem of experimental backgrounds. Early experiments employed the two-laser geometry of saturation spectroscopy (see Fig. 1), which involves looking for a small change in the intensity of a generally quite noisy laser beam. The use of the related two-laser geometry of polarization spectroscopy [5] (see Fig. 2) reduces the background due to the copropagating laser beam to that of polarizer leakage (six to eight orders of magnitude less), but introduces a new background—due to an induced dichroism effect. This new effect allows the determination of the material dephasing time, but its presence only complicates the method by causing several terms to occur in the expression to be curve fit. And if the dephasing time is an undesired quantity, these terms appear as a 100 percent shot-to-shot-jitter background acting to obscure the desired terms. The two-laser technique of Yajima and co-workers (see Fig. 3) avoids these backgrounds, but its unphase-matched nature requires the use of small angles between beams, and scattered light provides some background, although this method is probably the best two-laser arrangement. In general, two-laser geometries are constrained by the Bragg condition to necessarily have some background via the above mechanisms. Three-laser geometries, on the other hand, easily achieve background-free operation due to the extra free-

dom associated with the probe wavelength. See Fig. 5, for example, for one type of three-laser background-free arrangement. Others are not difficult to imagine.

Three-laser geometries also achieve simplified interpretation compared to that of two-laser arrangements in which the probe and excitation wavelengths are essentially equal. This equality in two-laser arrangements forces the probe beam to probe the ground-to-excitation-level transition, resulting in simultaneous probing of two induced gratings: a ground-state absorption grating and an excited-state stimulated-emission grating. This is in addition to any excited-state or intermediate-state absorption gratings that might be present also. The diffraction-efficiency line shape thus contains at least two and possibly more contributions, corresponding to two or more terms in the theoretical expressions for the line shape. Such additional terms complicate the interpretation of the data significantly. And since the data-interpretation and curve-fitting problems can be complex even for theoretical line shapes containing only a single term, it is important to avoid such unnecessary complications. The use of an independent probe wavelength allows the experimentalist to probe a unique system level, reducing the line shape to that due to a single level, if possible. In the language of four-wave mixing, the process becomes triply resonant with the third-order susceptibility containing only a single contribution. Thus, our conclusion is that the ease of data interpretation attained by using three input lasers justifies the extra experimental complexity and cost incurred in their use.

The use of three input beams raises additional questions regarding appropriate beam geometries that were already present for two-laser techniques, but which have now become more complex. The most important question is that of the "geometry bandwidth" or, equivalently, the avoidance of phase mismatch during wavelength tuning. A number of beam geometries that have proven very convenient for fixed-frequency experiments are unacceptable for variable-frequency experiments [13]. The phase-conjugate geometry, for example, has a minimal frequency bandwidth and should not be employed for experiments requiring large frequency scans. We have analyzed the problem of frequency bandwidths of N -wave-mixing beam geometries [13] and have shown that, in arrangements in which two frequencies vary (usually one input beam and the signal beam), to maximize the frequency bandwidth, the two variable-frequency beams should copropagate. And conversely, counterpropagation of these two beams results in minimal frequency bandwidth (e.g., the phase-conjugate geometry). In general, near copropagation of all beams best achieves the simultaneous goals of first phase matching the interaction and then maintaining phase matching (i.e., attaining large bandwidth). A counterpropagating-excitation-beam geometry (only one excitation beam varies in frequency) does prove convenient on occasion, however, and we have employed it for many of our experiments (see Fig. 10).

Another question that arises in tunable-laser-induced-grating experiments is that of the choice of experimental

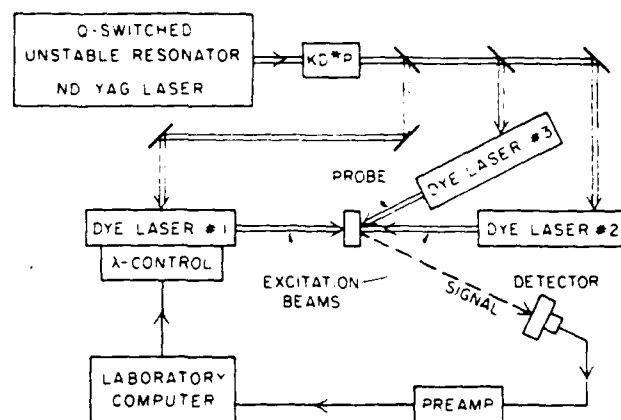


Fig. 10. Experimental apparatus employed in three-laser tunable-laser-induced-grating experiments by the authors.

parameters when some freedom of such choice exists. In experiments on organic dyes, for example, absorption bands are so broad that much freedom exists in the choice of excitation and probe wavelengths. In addition, the concentration, the beam intensities, and interaction length can also be free parameters in the experiment. While effects such as optical damage or limited laser power can act to constrain the input intensities, sample saturation may be the limiting factor. In this case, a simple calculation shows how to optimize the signal strength by appropriate choice of experimental parameters [7].

We can rewrite (16)–(18) to obtain an estimate of the magnitude of the diffraction efficiency in a typical experiment. If we define the saturation intensity in the usual manner ($I_{\text{sat}} = \hbar\omega_1/\sigma_{\text{ex}}\tau_0$), we have

$$\eta_i(0) = (I_1/I_{\text{sat}})(I_2/I_{\text{sat}})(N\sigma_{\text{ex}}L)^2(\sigma_{\text{pr}}/\sigma_{\text{ex}})^2(\tau_i/\tau_0)^2 \quad (27)$$

for $i = 0, 1, 2$. Equation (27) gives the peak diffraction efficiency in terms of dimensionless quantities that are often easy to evaluate for a typical experiment. In addition, the use of these dimensionless quantities allows us to incorporate important experiment constraints, facilitating the optimization of the diffraction efficiency for a given sample. For example, the two ratios I_1/I_{sat} and I_2/I_{sat} should be as large as possible to maximize the diffraction efficiency, but not as large as unity to avoid unwanted saturation effects. Typically, then, the values $I_1/I_{\text{sat}} = I_2/I_{\text{sat}} \approx 0.01$ – 0.1 will optimize the signal intensity without significantly sacrificing accuracy. Furthermore, the quantity $N\sigma_{\text{ex}}L$ will optimally assume a value on the order of unity to maximize the theoretical diffraction efficiency in the presence of beam attenuation due to absorption (which has not been included in the above expression). If an experiment employs the above values, i.e., $I_1/I_{\text{sat}} = I_2/I_{\text{sat}} = 0.1$, and $N\sigma_{\text{ex}}L = 1$, then

$$\eta_i(0) = 0.01(\sigma_{\text{pr}}/\sigma_{\text{ex}})^2(\tau_i/\tau_0)^2 \quad (28)$$

so that the peak diffraction efficiency is now written only in terms of physical parameters of the system. Interestingly, $\eta_i(0)$ is now *inversely* proportional to the square of the absorption cross section at the excitation wavelength

σ_{ex} [compare to (16)–(18)]. This results because a low value of σ_{ex} allows a long interaction length and a high excitation intensity (because I_{sat} will be large). Manipulation of the ratio σ_{pr}/σ_{ex} by a wise selection of experimental wavelengths can improve overall diffraction efficiency. And unlike the ratio τ_i/τ_0 which is necessarily less than one, the ratio of absorption cross sections can be large, thus leading to very high diffraction efficiencies. Of course, various additional factors, such as attenuation due to absorption and beam-intensity variations, will tend to reduce actual experimental efficiencies below the value given by (28); it is clear, however, that even in the presence of these effects, appropriate experimental parameters will result in detectable diffraction in most cases.

Having discussed the factors affecting the magnitude of tunable-laser-induced-grating effects, we now turn to a discussion of effects that alter the *shape* of the tunable-laser-induced-grating line shape from the ideal shape (26). Understanding these factors is critical for correct interpretation of experimental data.

The tunable-laser-induced-grating technique allows the determination of material relaxation times when significant line broadening, beyond mere lifetime broadening, is present. Because the result of an experiment is a curve whose width is given by the inverse lifetime, the tunable-laser-induced-grating technique can be regarded as a line-narrowing method, achieving narrowing all the way to the natural linewidth. Thus, the presence of broadening mechanisms such as collision broadening, variable molecular environments, and overlapping levels (such as occurs in organic dyes, for example) does not significantly affect the observed line shape.

The presence of Doppler broadening, however, can result in broadening of tunable-laser-induced-grating line shapes, depending on the laser beam geometry used [7]. To see this, observe that a molecule with velocity \vec{v} sees a Doppler-shifted excitation frequency $\omega'_i = \omega_i(1 - \vec{k}_i \cdot \vec{v}/c)$ where ω_i is the excitation frequency in the laboratory frame and \vec{k}_i is a unit vector in the direction \vec{k}_i . This molecule will then see a frequency difference $\Delta\omega' \approx \Delta\omega - 2\omega_i(v_{||}/c) \sin \theta$ where 2θ is the angle between the excitation beams and $v_{||}$ is the component of the molecule's velocity in the direction of the induced grating. Each molecule will contribute maximally to the measured line shape when it observes that $\Delta\omega' = 0$, which will occur for non-zero values of $\Delta\omega$ that depend on $v_{||}$. The tunable-laser-induced-grating line shape will be a superposition of individual line shapes centered at $\Delta\omega' = 0$, i.e., $\Delta\omega = 2\omega_i(v_{||}/c) \sin \theta$, and hence, will be broadened. If the Doppler linewidth is $\Delta\omega_{Doppler}$ and the tunable-laser-induced-grating linewidth in the absence of Doppler broadening is very small, then the actual measured linewidth will be

$$\Delta\omega_{HWHM} \approx \Delta\omega_{Doppler} \sin \theta. \quad (29)$$

Thus, copropagating excitation beams ($2\theta = 0^\circ$) result in no Doppler broadening, while counterpropagating excitation beams ($2\theta = 180^\circ$) yield a maximally broadening

line. The effects of Doppler broadening have been negligible compared to lifetime broadening in measurements on ultrafast systems, but if the method is employed to study longer relaxation times, Doppler broadening could force a copropagating-excitation-beam geometry.

Another line-shape-distorting effect results from the presence of coherent background. One source of such background is the nonresonant third-order susceptibility. Another is the resonant background due to any solvent or host material employed. While we have found both of the above effects negligible in most of our experiments, another effect—due to the choice of experimental wavelengths—has been responsible for a significant amount of line shape distortion and asymmetry in many of our experiments. Suppose that an experiment measures the third-order susceptibility $\chi_{ijk}^{(3)}(\omega_1 - \omega_2 + \omega_3)$ with ω_1 and ω_2 representing the excitation frequencies and ω_3 representing the probe frequency (see Table I). The phase-matching requirement will prevent all other sum- and difference-frequency processes from contributing radiation to the diffracted signal, except for the process corresponding to $\chi_{ilk}^{(3)}(\omega_3 - \omega_2 + \omega_1)$, which will necessarily contribute because its phase-matching requirement is identical to that of the former process. It corresponds to forming a grating with the frequencies ω_3 and ω_2 and probing it with ω_1 . If $|\omega_3 - \omega_2| \gg 1/\tau$, where τ is the relaxation time scale, then this latter term will be very small. If, on the other hand, this relation is not satisfied, the term $\chi_{ilk}^{(3)}(\omega_3 - \omega_2 + \omega_1)$ will contribute coherent background, which will produce noticeable distortion in the line shape. In our experiments on organic dyes, this term could be accurately approximated by a constant complex quantity $C_1 + iC_2$ to be found by the curve-fitting program. Alternatively, precise knowledge of $\omega_3 - \omega_2$ would allow the curve-fitting program to calculate $C_1 + iC_2$ for any potential theoretical line shape function, thus adding no additional parameters to the fitting process.

A good curve-fitting procedure is essential to data interpretation in tunable-laser-induced-grating experiments. It is not possible to inverse-Fourier-transform the experimental line shape because of the potential coherent backgrounds, but more importantly, also because the line shape is the *squared* modulus of the Fourier transform of the material response. We employ weighted least squares using a simplex minimization routine, although a Newton-Marquardt technique is much faster but more complex [51], [52]. Goodness-of-fit tests are essential [53]; we have found that the "runs test" and the "Durbin-Watson test" are the most powerful tests of a fit; nevertheless, we also employ the "Kolmogorov-Smirnov test" and the "runs-up-runs-down test." The difficulty of accurately measuring the absolute experimental error prevents the well-known " χ^2 test" from providing much useful information in tunable-laser-induced-grating experiments. In addition, an F test is necessary to determine whether a complex theoretical model is needed over a simpler one [53], e.g., whether a sum-of-exponentials model is needed over a single exponential decay. We fit

TABLE II
SOME DECAY LAWS AND THEIR TUNABLE-LASER-INDUCED-GRATING LINE SHAPES*

	Decay Function $h(t)$	Line Shape $R(\Delta\omega) + iI(\Delta\omega)$
Exponential	$\exp(-t/\tau)$	$1/(1 + i\Delta\omega\tau)$
Sum of Exponentials	$A \exp(-t/\tau_1) + (1 - A) \exp(-t/\tau_2)$	$A\tau_1/(1 + i\Delta\omega\tau_1) + (1 - A)/(1 + i\Delta\omega\tau_2)$
Range of Exponentials	$\int_{\tau_1}^{\tau_2} \exp(-t/\tau) d(1/\tau)$ $= [\exp(-t/\tau_1) - \exp(-t/\tau_2)]/t$	$(1/2) \log_e \left[\frac{1 + (1/\Delta\omega\tau_2)^2}{1 + (1/\Delta\omega\tau_1)^2} \right]$ $+ i \text{Arctan}[1/\Delta\omega\tau_1] - i \text{Arctan}[1/\Delta\omega\tau_2]$

*Normalization constants (which yield $h(0) = R(0) = 1$) have been omitted in the range-of-exponentials functions for brevity.

our data to the following expression:

$$\eta(\Delta\omega) = C_0 + H \cdot \{[R(\Delta\omega) + C_1]^2 + [I(\Delta\omega) + C_2]^2\} \cdot \{1 + \beta \cdot \Delta\omega\} \quad (30)$$

where $R(\Delta\omega)$ and $I(\Delta\omega)$ represent the real and imaginary parts, respectively, of the Fourier transform of the material impulse response; C_0 represents all incoherent backgrounds and electronic baselines; H incorporates the line strength and detector responsivities; C_1 and C_2 represent any coherent background that may be present; and β represents a variety of wavelength-dependent effects including material absorption, dye-laser pulse length and timing, and detector responsivity. Our experience has been that $\beta \approx 0$ in all experiments on dyes and optical-Kerr materials, which have lifetimes as short as ≈ 100 fs. For typical line shape functions $R(\Delta\omega) + iI(\Delta\omega)$, see Table II, which lists some theoretical decays and their corresponding line shapes. Finally, in nearly all of our experiments, we are able to replace C_1 and C_2 by the quantities $R(\omega_3 - \omega_1)$ and $I(\omega_3 - \omega_1)$, respectively, thus reducing the number of variables to be fit from eight to seven or six (depending on whether $\omega_3 - \omega_1$ is accurately known or needs to be fit). In any case, satisfactory fits are always obtained, subject to the quality of the theoretical model.

V. EXPERIMENTAL APPARATUS

Fig. 10 gives a schematic of the experimental apparatus that we have used to perform tunable-laser-induced-grating experiments. A homemade Q -switched, unstable-resonator Nd:YAG laser provides 10 ns, ≈ 120 mJ pulses at 1.064 μm at a repetition rate of 10 pulses/s. A temperature-stabilized CD*A or KD*P crystal frequency doubles this radiation. Beam splitters then divide 20–30 mJ of the 532 nm radiation into three beams of equal energy to pump three dye lasers of prism-beam-expander/grazing-incidence hybrid design [54]. Using the dyes rhodamine 590 or 640, each dye laser provides 7 ns pulses of between 200 and 800 μJ with line widths between 0.1 and 0.4 cm^{-1} .

The dye-laser beams are collimated, apertured, and after considerable beam steering, focused with AR-coated 15–35 mm focal-length lenses into a 1 mm thick unstirred glass sample cell. Spot sizes inside the sample cell are approximately 30–80 μm , and peak intensities are about 0.1–3 GW/cm^2 . The excitation beams propagate in a pre-

cisely counterpropagating geometry, with the probe beam at an angle of several degrees with respect to the fixed-frequency excitation beam. Polarizers placed adjacent to the sample cell ensure that each input beam and the signal beam are highly polarized in the appropriate direction. Spatial filtering of the signal beam at the detector is achieved with an aperture/lens/focal-plane-aperture combination, and wavelength filtering employs a Schott OG550 long-pass filter.

In all experiments, the signal beam is detected by a PIN-8LC United Detector Technology silicon detector whose output is preamplified by factors of 5–200 to achieve signal levels of 100 mV–1 V. All voltage signals are deliberately stretched in time to ≈ 50 μs pulse lengths so that the required electronic bandwidths are minimal. Sample-and-hold circuits capture the preamp output and, in addition, the outputs of detectors monitoring all three dye-laser outputs. A DEC MINC (LSI-11) laboratory computer reads all four of these voltages, subtracts off any background (determined with a four-phase laser-shutter scheme), and normalizes the signal by the product of the three dye-laser energies. The MINC also scans the variable-frequency dye laser.

Experimental runs generally consist of 1–20 wavelength scans, each containing 40–120 data points, with each data point the result of 100–1000 laser shots. The shot-to-shot jitter in the normalized signal is also measured, and the final signal level is defined to be the error-weighted average over all scans. Experimental runs require from 30 min to 8 h. Finally, it is important to perform additional runs at different intensities to verify the linearity of results, thus eliminating the possibility of saturation-induced lineshape distortion.

VI. ORGANIC-DYE EXPERIMENTS

Nonradiative relaxation in nonfluorescent dyes has long been a subject of great interest, both for its own inherent scientific value [55] and for its various applications, e.g., the mode locking of lasers [56]. The class of triphenyl methane dyes is a particularly interesting example of such dyes, exhibiting a strongly solvent-viscosity-dependent relaxation time scale over several orders of magnitude of variation of the solvent viscosity [57]. While a detailed discussion of the complex dynamics of this class of molecules is far beyond the scope of this work (see, for example, the papers by Sundstrom and co-workers [58],

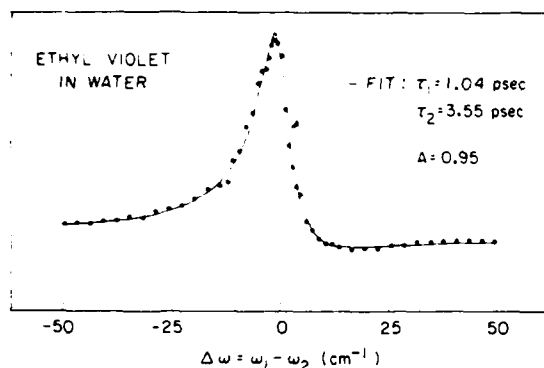


Fig. 11. Experimental line shape for ethyl violet in water.

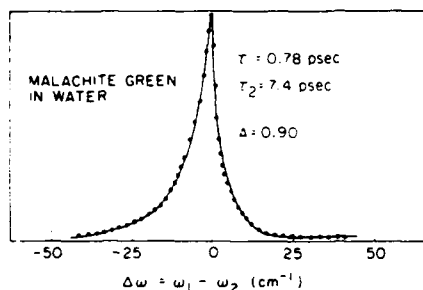


Fig. 12. Experimental line shape for malachite green in water.

[59]), we mention here that for low-viscosity solvents, the excited-state decay in triphenyl methane dyes is rate limited by internal conversion from the S_1 to S_0 levels. The relaxation time scale for this process can be as short as a fraction of a picosecond, and as a result, researchers have employed both ultrashort pulse time-domain techniques [60] and tunable-laser-induced-grating frequency-domain techniques [10] to study these molecules.

We will discuss the results of some of our frequency-domain experiments on these dyes; more complete experiments have been performed on this class of molecules in the time domain using ultrashort pulses [58], [59]. Our purpose here, however, is merely to illustrate the utility of frequency-domain techniques for the measurement of ultrafast excited-state decay in such molecules.

The tunable-laser-induced-grating experimental line shapes for the triphenyl methane dyes, ethyl violet, and malachite green, each dissolved in water at room temperature, are shown in Figs. 11 and 12. In each case, the wavelengths were chosen to ensure that the experiment probed the minimum number of energy levels to minimize the complexity of the data-interpretation problem. Specifically, the probe wavelength was set slightly smaller than that of the excitation beams to attempt to probe only the ground level. In this way, we can study the *ground-state recovery* with minimal additional effects also contributing to the experimental line shape.

Our data for ethyl violet (excited at 570 nm and probed at 569.5 nm), with a half-width half-maximum of 2.9 cm^{-1} , indicate a ground-state-recovery time scale of about 1.2 ps. Indeed, curve fitting the data in Fig. 11 to a Lorentzian (with coherent background; see Section IV) yields

a fit with this lifetime. The solid line through the data shows this fit, whose relaxation time scale is close to that observed by Sundström *et al.* [58], [59] for similar-viscosity solvents. A fit to a more complex line shape, the sum of two Lorentzians, corresponding to a decay consisting of the sum of two exponentials, yields a slightly better fit, however. Although it is visually indistinguishable from the simpler single exponential, the performance of an F test indicates that the biexponential decay containing a small component of a slightly longer relaxation is a significantly better fit to the data. The biexponential fit is $h(t) = A \cdot \exp(-t/\tau_1) + (1 - A) \cdot \exp(-t/\tau_2)$ with $A = 0.95$, $\tau_1 = 1.04 \text{ ps}$, and $\tau_2 = 3.55 \text{ ps}$; the significance level, given by the F test, is 2.5 percent. Additional experiments are necessary to verify this decay.

Biexponential behavior is also observed for malachite green in water (excited at 613.4 nm and probed at 609.5 nm). In this case, however, the difference between the two fits is quite striking: the single-Lorentzian line shape is a visually very bad fit to the data, while the sum of two exponentials provides a quite good fit. The solid line through the data in Fig. 12 indicates the latter fit, corresponding to the decay $h(t) = A \cdot \exp(-t/\tau_1) + (1 - A) \cdot \exp(-t/\tau_2)$ with $A = 0.9$, $\tau_1 = 0.78 \pm .1 \text{ ps}$, and $\tau_2 = 7.4 \pm 3 \text{ ps}$, which is in approximate agreement with the ultrashort-pulse experiment of Migus *et al.* [61] who also observed a biexponential decay, although with somewhat shorter time scales.

The existence of two exponentials in the ground-state recovery of ethyl violet and malachite green could be due to any of several possible causes. First, two ground-state conformers could be present simultaneously in solution, with each having its own single-exponential decay [58], [59]. Second, the decay is not a single-step process: in addition to the internal conversion, two extremely fast vibrational relaxations must occur for ground-state recovery to occur. Such a multistep process could produce a line shape that could resemble a multiexponential decay. This cause is doubtful for these dyes because these latter decays are expected to be at least an order of magnitude faster than the internal conversion. Other levels, such as that of a different molecular conformer created as a result of the excitation, could, however, be the cause [59]. Third, despite our attempts to probe only a single level, it is possible that population of some intermediate level is occurring, leading to the probing of an additional level. In this case, the diffracted fields would add coherently, and the resulting line shape would be the sum of two exponentials. And finally, the actual line shape could be more complex than sums of exponentials, with the sum of two or more exponentials merely providing a convenient approximation of the actual line shape. For example, we curve-fit the malachite green data to the line shape corresponding to a range of simple exponentials [62] (see Table II) and obtained a fit that was indistinguishable from that obtained with the sum of two exponentials. The minimum and maximum lifetimes were 0.34 and 7.2 ps, respectively. *What is important to realize is that the com-*

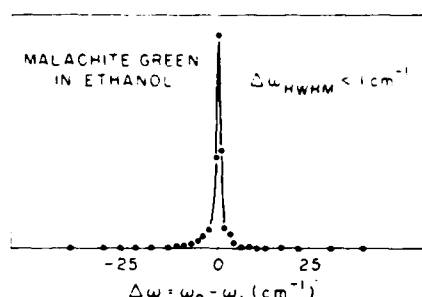


Fig. 13. Experimental line shape for malachite green in ethanol using parallel-polarized excitation beams. These data represent a thermal grating, and not an excited-state grating.

plex line shapes observed here reflect the physics of the molecules rather than artifacts of the technique. These results are typical of the data that have been obtained in our tunable-laser-induced-grating studies on absorbing samples, which have shed some interesting light on the ultrafast excited-state decay in triphenyl methane dyes. While many other experiments have dealt with the problem of internal conversion in these molecules, much more remains to be done. It is clear, however, that tunable-laser-induced-grating techniques provide a useful and convenient method for such study.

VII. THERMAL GRATINGS AND THEIR SUPPRESSION

Both of the experiments described in the previous section involved the use of water as the solvent. Replacing water as the solvent in tunable-laser-induced-grating malachite green experiments with ethanol, methanol, or chloroform produces drastic changes in the observed line shape (see Fig. 13). The line width decreases by about an order of magnitude to approximately the laser line width, and the peak diffraction efficiency increases by more than an order of magnitude. Both two- and three-laser experiments exhibit these effects. An increase in the internal conversion time in these solvents by a factor of ten or so would be consistent with these observations, but time-domain experiments [57] indicate an approximately 2 ps ground-state-recovery time for malachite green in ethanol as well as in water. The actual cause of this observed narrowing lies in the formation of a *thermal grating* [41].

A thermal grating forms when radiation is absorbed from crossed beams in a nonradiatively decaying material, and a sinusoidal temperature distribution results. This spatial temperature modulation or "thermal grating" will diffract the light of a probe beam because, in general, a medium's refractive index varies with temperature. Such thermally induced phase gratings can be quite efficient, and as a result, diffracted light from a thermal grating can obscure the light from a population grating and seriously complicate the determination of excited-state lifetimes. This complicating effect takes the form of a narrow spike centered at $\Delta\omega = 0$ because thermal gratings decay by thermal diffusion, a relatively slow process.

In general, the thermal-grating line shape can be determined by the general Fourier-transform arguments of Section III-C with the knowledge that the decay function is

given by an exponential with time constant $\tau = (k_{gr}^2 D)^{-1}$ where k_{gr} is the magnitude of the material grating k vector and D is the thermal diffusivity [63]. This time scale is typically much longer than the laser pulses used in tunable-laser-induced-grating experiments. As a result, the CW theory of Section III-C breaks down, and thermal-grating formation must be treated as an integrated-intensity effect [64].

Integrated-intensity gratings yield an interesting class of problems in which the partial-coherence properties of the excitation beams play a key role. Their theoretical treatment, however, is beyond the scope of this work. See [41] and [64] for more information on this subject. Our interest here is, rather, in the *suppression* of this effect, since it prevents the measurement of the desired population relaxation.

To appreciate the difficulty of the problem of thermal-grating suppression, first observe that, theoretically, both the population-grating strength and the thermal-grating strength depend similarly on nearly all of the various experimental quantities, such as concentration, intensity, excitation absorption cross section, and interaction length. And, in addition, thermal-grating effects are quite strong, in general, so that much suppression is necessary, i.e., two or more orders of magnitude, depending on the material.

Various suggestions for thermal-grating suppression have appeared in the literature, most of which have allowed excited-state-lifetime measurement in a few cases where, previously, thermal gratings have dominated. Saitkan and Sei [65] observed that, contrary to expectation, at very weak concentrations, the ratio of population- to thermal-grating strength increases. Desai *et al.* [66] calculated that a nonzero hydrodynamic thermal-grating buildup time necessarily exists, and using picosecond pulses in conjunction with large fringe spacing could suppress thermal gratings by a large factor. This method remains to be tried, however. Trebino *et al.* [67] showed that some suppression could be achieved by using optical homodyne detection with partially coherent light using the polarization spectroscopy geometry. Suppression using this method was, however, not sufficient to merit its experimental difficulties: pulsed dye lasers are, in general, too noisy a source to perform optical homodyne detection with.

The best thermal-grating suppression method suggested to date is to use *orthogonally polarized excitation beams* [68]. The use of orthogonally polarized excitation beams to suppress grating effects has been known for a century, and time-domain ultrashort-pulse techniques often employ orthogonal polarizations to suppress the "coherence spike" [12]. The use of such an arrangement in a technique that depends intimately on the formation of induced gratings requires the more recent notion that orthogonally polarized excitation beams, while unable to induce scalar gratings, can, in fact, induce gratings—*tensor* gratings. Since excited-state gratings are molecular-polarizability (i.e., tensor) gratings and thermal gratings act through

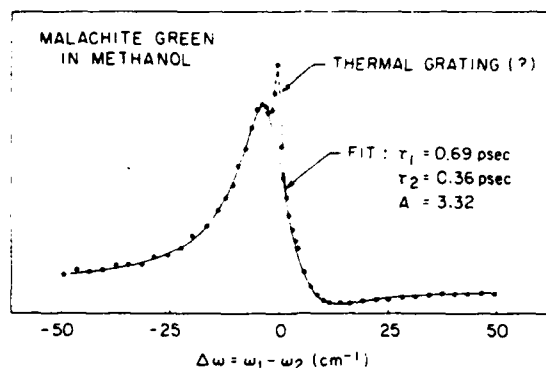


Fig. 14. Experimental line shape for malachite green in methanol using orthogonally polarized excitation beams. Observe that only a small remnant of the thermal grating remains.

temperature (i.e., scalar) gratings, orthogonally polarized excitation beams produce excited-state gratings, but not thermal gratings. Suppression is, in theory, perfect.

In experiments on malachite green dissolved in methanol, using orthogonally polarized excitation beams, thermal-grating suppression by a large factor is observed (see Fig. 14). While previous experiments using parallel polarized excitation beams saw mainly the thermal gratings, these experiments see mainly population-grating effects, allowing the extraction of a biexponential decay law without difficulty.

It is interesting that in the above experiment, some thermal-grating component appears to remain, forcing the rejection of several data points near line center. The most likely explanation for this probable thermal-grating remnant is that, since the probe beam necessarily has the same polarization as one of the excitation beams, it acts as an excitation beam in conjunction with the excitation beam of the same polarization to cause a small component of temperature modulation to occur. This presumably occurs because, in the particular experiment above, the probe wavelength is not far from the excitation wavelengths; the thermal-grating remnant would thus have the same origin here does the asymmetry (see Section IV). Separating the probe wavelength from the excitation wavelengths would then remove both the asymmetry and the thermal-grating remnant. Such experiments are underway at the moment.

In any case, it is clear that thermal gratings can be suppressed sufficiently to perform high-quality measurements on absorbing samples.

VIII. OPTICAL-KERR-EFFECT EXPERIMENTS

Since its first experimental observation in 1964 [69], the optical Kerr effect has been of great interest in nonlinear optical research, achieving application to phase-conjugate reflection [70], picosecond optical gating [71], optical bistability [72], and a wide variety of other techniques. The physical effect responsible for these applications is summarized by the relation (valid in the CW limit)

$$n_i = n_0 + n_2 \cdot I, \quad (31)$$

that is, the index of refraction for light polarized in the i

direction depends on the intensity of light present that is polarized in that direction. Two similarly polarized beams that interfere in an optical-Kerr medium thus induce a phase grating in that medium.

The physical mechanism underlying the optical Kerr effect is the reorientation of molecules by the applied optical electric field [73]. More precisely, the applied field induces in an ordinarily isotropic medium an anisotropic molecular and/or electronic distribution function. Initially, randomly oriented molecules incur a dipole moment due to the applied field. These induced dipoles then rotate slightly in the applied field producing anisotropy whose preferred direction is parallel to the applied electric field polarization. The relaxation of such an anisotropy is then the *orientational* relaxation of molecules back to the randomly oriented, isotropic distribution.

Knowledge of the orientational relaxation of the optically induced anisotropy in an optical-Kerr liquid is important for understanding the dynamics of molecules in the liquid state. As a result, numerous experiments have measured orientational relaxation in optical-Kerr liquids. Depolarized light scattering measures, in the frequency domain, one component of the third-order susceptibility tensor $\chi_{xxx}^{(3)}$ [73] (corresponding to the formation of a grating with orthogonally polarized excitation beams), and a large number of experimental light scattering studies have been performed [74], [75]. Data for the liquid CS_2 show at least two components, corresponding to a relatively long "Debye" component and a shorter, "interaction-induced" component [73], [75]. At room temperature, the Debye component line shape fits quite well to a Lorentzian with a half-width half maximum of 3.5 cm^{-1} , corresponding to an exponential decay with a time constant of 1.5 ps, in rough agreement with orientational-diffusion calculations [73]. Deviations from a Lorentzian line shape in the wings of light-scattering data indicate the presence of the faster interaction-induced component. Although its precise shape is difficult to determine from the data due to the effects of cross terms, it is clearly not Lorentzian, that is, this component of the decay is not exponential [73]. More precisely, the overall decay—and we must use this term loosely because more than one process may be contributing—appears not to be the sum of two exponentials. Nonetheless, if we approximate the decay by the sum of two exponentials, the measured time scale for the fast or interaction-induced component of the decay is approximately 0.2 ps.

Recent ultrashort-pulse experiments have at least partially confirmed these observations. Using an interferometric technique with 70 fs pulses, Halbout and Tang [76] have observed the decay of the $\chi_{xxx}^{(3)}$ component of the third-order susceptibility for the optical Kerr effect in CS_2 . Lacking the temporal resolution to study the actual shape of the fast component of the decay, they fit their data to the empirical curve consisting of the sum of two exponentials, obtaining the best fit with the decay $h(t) = A \cdot \exp(-t/\tau_1) + (1 - A) \cdot \exp(-t/\tau_2)$ with $A = 0.14$, $\tau_1 = 2.00 \text{ ps}$, and $\tau_2 = 0.33 \text{ ps}$, in rough agreement with

light-scattering results, if the different tensor elements measured by these two types of experiments are expected to yield the same time scales. Another time-domain experiment was performed by Greene and Farrow [77] who used an induced-dichroism method and 150 fs pulses and measured the quantity $\chi_{xxv}^{(3)} + \chi_{vvx}^{(3)}$. Fitting to the same empirical curve, they obtained a best fit with the parameters $\tau_1 = 2.16 \pm 0.1$ ps and $\tau_2 = 0.24 \pm .02$ ps. Neither of these two experiments was able to observe deviations from an exponential decay at short time scales due to the finite-pulse-length effects inherent in their experiments.

A number of other techniques have addressed the problem of orientational relaxation of molecules. These techniques include Raman scattering, neutron scattering, infrared absorption, dielectric relaxation, nuclear magnetic resonance, and fluorescence. A discussion of these additional techniques is beyond the scope of this paper; the interested reader is referred to Berne and Pecora's work [78] for an introduction to these methods and a comparison of their measured quantities.

We elected to study the optical Kerr effect in CS₂ using the tunable-laser-induced-grating technique because, unlike light-scattering experiments, it has the capability to measure *all* of the available third-order susceptibility elements. And unlike ultrashort-pulse techniques, temporal resolution is not limited by pulse length effects (in addition, pulse-spreading effects in the sample are not a problem).

Using the previously described apparatus with an excitation wavelength of 575 nm and a probe wavelength of 570 nm, we studied CS₂ at room temperature and measured all four nonzero, distinct third-order-susceptibility elements. High-quality polarizers were used in all four beams to ensure that the measurement of one element was not contaminated by another. Fig. 14 shows the data (observe the symmetry of the line shapes due to the increased difference in frequency between excitation and probe frequencies compared to the data for the triphenyl methane dyes). We again curve fit the data to line shapes consisting of the sum of two Lorentzians, and Table III summarizes the best fit parameters for the four curves with rough estimates for the errors. As the data in Table III indicate, we observe both components of the Kerr effect.

While the lack of absolute measurements makes it difficult to relate these results to the experiment of Greene and Farrow, we can compare them to light-scattering measurements and the experiment of Halbout and Tang. Our values for the relaxation time scales of $\chi_{xxv}^{(3)}$ of 1.26 and 0.13 ps are in approximate agreement with light-scattering data. On the other hand, light scattering experiments yield a value of A of about 0.55 [73], somewhat different from our experimental value of 0.23.

Our values for the relaxation timescales of $\chi_{xxx}^{(3)}$ of 2.49 and 0.21 ps should be compared to the ultrashort-pulse values of Halbout and Tang, who measured 2.00 and 0.33 ps for these timescales, respectively. Their value of A , 0.14, is very close to ours, which is 0.13.

Unfortunately, to our knowledge, no other independent

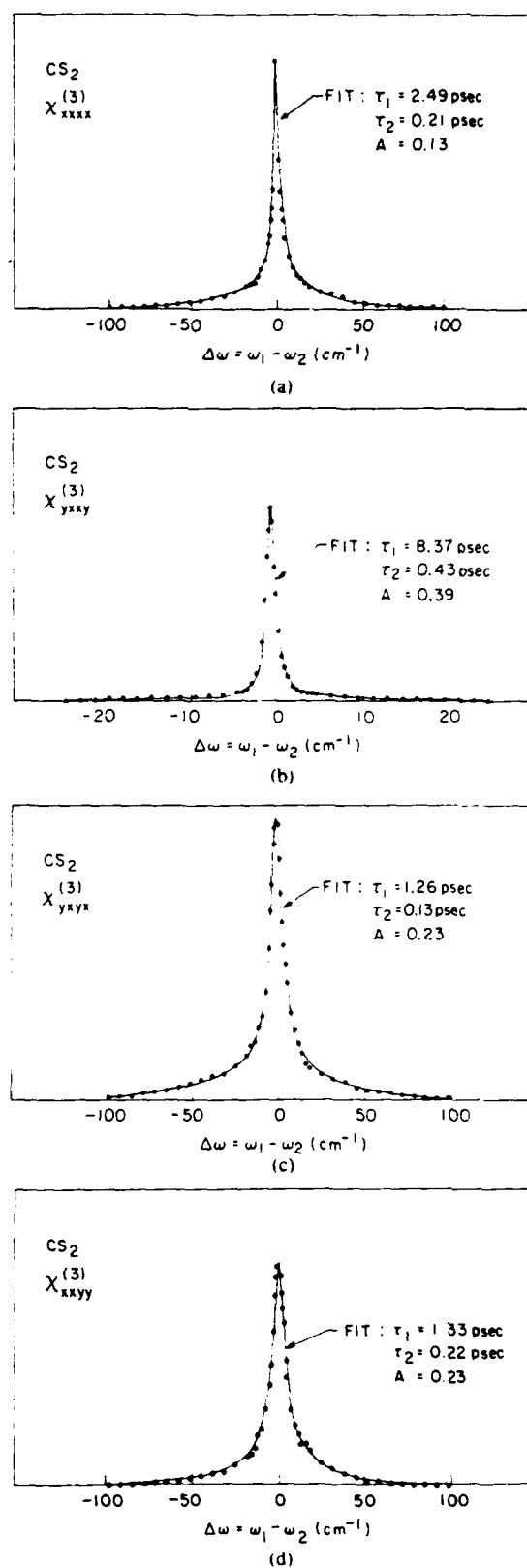


Fig. 15. Experimental line shapes for the four distinct, nonzero third-order-susceptibility tensor elements of the optical Kerr effect in carbon disulfide.

measurement of the line shape or decay of $\chi_{xxx}^{(3)}$ exists to which we may compare our results.

We were, however able to perform a partial check on

TABLE III
EXPERIMENTAL VALUES FOR THE RELAXATION TIME SCALES OF
CARBON DISULFIDE AS MEASURED BY THE TUNABLE-LASER-
INDUCED-GRATING TECHNIQUE*

Tensor Element	Slow Lifetime (ps)	Fast Lifetime (ps)	Weight (A)
$\chi_{xxxx}^{(3)}$	2.49	0.21	0.13
$\chi_{xyxy}^{(3)}$	8.37	0.43	0.39
$\chi_{yyxx}^{(3)}$	1.26	0.13	0.23
$\chi_{xyyx}^{(3)}$	1.33	0.22	0.23

*The line shape data were fit to the squared modulus of the Fourier transform of the decay given by the theoretical expression $h(t) = A \exp(-t/\tau_1) + (1 - A) \exp(-t/\tau_2)$, where τ_1 is the "slow" lifetime, τ_2 is the fast lifetime, and A is the weight of the slow component. Error on each quantity are estimated to be ± 20 percent.

our results by measuring the fourth tensor element $\chi_{xyxy}^{(4)}$ which should in theory [79] equal $\chi_{xyxy}^{(3)}$. Our values for $\chi_{xyxy}^{(3)}$ are $\tau_1 = 1.33$ ps, $\tau_2 = 0.22$ ps, and $A = 0.23$; these values should be compared to our values of $\tau_1 = 1.26$ ps, $\tau_2 = 0.13$ ps, and $A = 0.23$ for $\chi_{yyxx}^{(3)}$. Since the approximate errors in these parameters are about ± 0.2 for τ_1 , ± 0.05 for τ_2 , and ± 0.05 for A , this agreement is quite good.

It is not well understood at this time why such a variation of lifetimes should occur in these elements, and in particular, why the line shapes obtained with orthogonally polarized excitation beams should agree with each other but not with line shapes obtained with parallel-polarized excitation beams (which themselves do not agree with each other). Simple classical orientational-diffusion models predict the same decays for all four tensor elements [78]. By performing attenuation experiments, we have verified that the diffracted beam energy varies linearly with the three input beam energies over more than two orders of magnitude of input energy; we can thus eliminate saturation effects from consideration. We have reperformed the experiments with ω_3 a few angstroms away from ω_1 so that the resulting severe asymmetry in the line shape would reveal a potential narrow incoherent spike at $\Delta\omega = 0$ (as in Fig. 14); we saw no such component. (Such a component could come about due, for example, to a thermal grating produced by small amounts of absorption in the sample. If present, it would only appear in $\chi_{xxxx}^{(3)}$ and $\chi_{xyxy}^{(3)}$ and it would cause narrower line shapes in those elements, resulting in the appearance of longer lifetimes in these elements.) In addition, we have repeated the experiments and have obtained similar results.

One possible explanation of the above results comes from nonpolarized light-scattering data in which spontaneous Brillouin scattering contributes a "doublet" of spectral lines at $\pm\omega_B$ where $\omega_B/v_{\text{acoustic}}$ is the acoustic wave vector magnitude that conserves momentum in the interaction and v_{acoustic} is the velocity of sound in the medium. In our parallel-polarized-excitation-beam experiments, it is likely that light-induced acoustic waves with a k vector

equal in magnitude to that of the light-induced grating will counterpropagate throughout the interaction region due to the sinusoidal of the intensity [80]. In orthogonally polarized-excitation-beam experiments, on the other hand, the uniform intensity in the medium should prevent such effects (which occur via an $E \cdot \nabla E = \frac{1}{2} \nabla(E^2)$ term). Thus, "Brillouin" lines would occur in $\chi_{xxxx}^{(3)}$ and $\chi_{xyxy}^{(3)}$, but not in $\chi_{xyyx}^{(3)}$ or $\chi_{yyxx}^{(3)}$. When they occur, their widths would be small, and their spectral positions would be at ± 0.22 cm^{-1} for CS_2 in the above geometry—too closely spaced to be resolved in our experiments, but at the same time, sufficiently separated to appear nonartificial, i.e., sufficiently spaced to result in a spectral feature significantly broader than the laser line widths. If this is the case, then the above measured line shapes for $\chi_{xxxx}^{(3)}$ and $\chi_{xyxy}^{(3)}$ would appear anomalously narrow and would curve fit to lifetimes that are anomalously long.

This explanation would indicate that, as in tunable-laser-induced-grating experiments on absorbing samples in which thermal gratings obscure lifetime information in parallel-polarized-excitation-beam arrangements, somewhat analogous effects could obscure line shapes in parallel-polarized-excitation-beam experiments on nonabsorbing samples. Experiments underway at the moment are aimed at clarifying this issue. It is clear, however, that in all tunable-laser-induced-grating experiments, parallel-polarized-excitation-beam arrangements are subject to intensity-dependent effects that can complicate data interpretation significantly, while experiments using orthogonally polarized excitation beams appear to date to be artifact free. *Thus, unless these intensity-dependent effects are desired, the use of orthogonally polarized excitation beams is, in general, advised for tunable-laser-induced-grating experiments on ultrafast phenomena.*

IX. CONCLUSIONS AND FUTURE DIRECTIONS

Tunable-laser-induced-grating techniques have proven quite effective for the study of ultrafast phenomena. Effects that have in the past complicated these methods are now well understood and can be eliminated by proper choice of experimental parameters. As a result, experimental data have improved significantly in the past decade. In addition, experimental interpretation has also become relatively straightforward.

Studies to date have obtained excellent results on both liquids and solids in the 100 fs–10 ps range. Current frequency-domain efforts are extending the upper endpoint of this range to values in excess of 1 s [81], while future studies should extend the lower endpoint of this range to values below 10 fs. Ultrafast processes that will, as a result, come under study with dynamical-grating techniques will include such processes as electron-phonon interactions in semiconductors, dephasing in liquids, electron solvation in liquids, and vibrational relaxation within an electronic manifold in organic dyes in solution (to name a few), all occurring on time scales as short as 10^{-14} s. It is evident that tunable-laser-induced-grating techniques for ultrafast-event measurement have reached a quite ma-

ture stage: the era of technique development is drawing to a close, and future work will focus on materials studies in regimes where such studies will be essential to our understanding of fundamental states of matter.

ACKNOWLEDGMENT

The authors would like to thank G. Kenney-Wallace for a critical reading of the manuscript and for her always-welcome advice. We would also like to acknowledge the skillful assistance of D. Arnone for his mechanical help in these experiments. In addition, we are indebted to A. Kostenbauder and B. Yoo for numerous helpful conversations.

REFERENCES

- [1] A. M. Weiner, J. G. Fujimoto, and E. P. Ippen, "Compression and shaping of femtosecond pulses," in *Ultrafast Phenomena IV*, Berlin: Springer-Verlag, 1984, pp. 11-15.
- [2] M. C. Downer, R. L. Fork, and C. V. Shank, "Imaging with femtosecond optical pulses," in *Ultrafast Phenomena IV*, Berlin: Springer-Verlag, 1984, pp. 106-110.
- [3] T. Yajima, "Non-linear optical spectroscopy of an inhomogeneously broadened resonant transition by means of three wave mixing," *Opt. Commun.*, vol. 14, pp. 378-382, July 1975.
- [4] F. Keilmann, "Tunable-laser-induced-grating dip for measuring subpicosecond relaxation," *Appl. Phys.*, vol. 14, pp. 29-33, 1977.
- [5] J. J. Song, J. H. Lee, and M. D. Levenson, "Picosecond relaxation measurements by polarization spectroscopy in condensed phases," *Phys. Rev. A*, vol. 17, no. 4, pp. 1439-1447, 1978.
- [6] A. E. Siegman, "Proposed picosecond excited-state measurement method using a tunable-laser-induced-grating method," *Appl. Phys. Lett.*, vol. 30, no. 1, pp. 21-23, 1977.
- [7] R. Trebino, "Subpicosecond-relaxation studies using tunable-laser-induced-grating techniques," Ph.D. dissertation, Stanford Univ., Stanford, CA, May 1983.
- [8] T. Yajima, H. Souma, and Y. Ishida, "Study of ultrafast relaxation processes by resonant Rayleigh-type optical mixing. II. Experiment on dye solutions," *Phys. Rev. A*, vol. 17, no. 1, pp. 324-334, 1978.
- [9] F. Keilmann, "Infrared saturation spectroscopy in p-type germanium," *IEEE J. Quantum Electron.*, vol. QE-12, pp. 592-597, Oct. 1976.
- [10] R. Trebino and A. E. Siegman, "Subpicosecond relaxation study of malachite green using a three-laser frequency-domain technique," *J. Chem. Phys.*, vol. 79, no. 8, pp. 3621-3626, 1983.
- [11] T. Yajima and H. Souma, "Study of ultrafast relaxation processes by resonant Rayleigh-type optical mixing. I. Theory," *Phys. Rev. A*, vol. 17, no. 1, pp. 309-323, 1978.
- [12] D. Kuhlke, "Transient orientational grating technique for investigations of fast molecular relaxation processes," *Appl. Phys. B*, vol. 34, pp. 129-137, 1984.
- [13] R. Trebino and A. E. Siegman, "Frequency bandwidths in nondegenerate *N*-wave-mixing interactions and induced-grating geometries," *Opt. Commun.*, vol. 56, no. 4, pp. 297-302, Dec. 15, 1985.
- [14] J. F. Lam, D. G. Steel, and R. A. McFarlane, "Collisionally induced narrowing of the longitudinal relaxation linewidth in nearly degenerate four-wave mixing," *Phys. Rev. Lett.*, vol. 49, pp. 1628-1631, Nov. 29, 1982.
- [15] W. E. Lamb, Jr., "Theory of an optical maser," *Phys. Rev.*, vol. 134, pp. A1429-A1450, June 15, 1964.
- [16] E. V. Baklanov and V. P. Chebotayev, "Effects of Fields in resonant interaction of opposing waves in a gas. I," *Sov. Phys. JETP*, vol. 33, pp. 300-308, Aug. 1971.
- [17] —, "Resonance interaction of unidirectional waves in gases," *Sov. Phys. JETP*, vol. 34, pp. 490-494, Mar. 1972.
- [18] M. Sargent III, "Laser saturation grating phenomena," *Appl. Phys.*, vol. 9, pp. 127-141, 1972.
- [19] M. Sargent III, P. E. Toschek, and H.-G. Danielmeyer, "Unidirectional saturation spectroscopy. I. Theory and short-dipole-lifetime limit," *Appl. Phys.*, vol. 11, pp. 55-62, 1976.
- [20] M. Sargent III and P. E. Toschek, "Unidirectional saturation spectroscopy. II: General lifetimes, interpretations, and analogies," *Appl. Phys.*, vol. 11, pp. 107-120, 1976.
- [21] F. Keilmann, "Hot-hole relaxation in p-Ge investigated by saturation spectroscopy," *Il Nuovo Cimento*, vol. 39B, pp. 666-670, June 11, 1977.
- [22] F. Keilmann and J. Kuhl, "Broadband modulation of 3- μ m light by 10 μ m light in p-germanium," *IEEE J. Quantum Electron.*, vol. QE-14, pp. 203-206, Mar. 1976.
- [23] F. Keilmann, "Test of electronic band structure by laser-induced resonance," *Solid State Commun.*, vol. 25, no. 7, pp. 451-453, 1978.
- [24] T. Grave and F. Keilmann, "Relaxation dynamics and photoconductivity in p-type germanium," *Z. Phys. B*, vol. 32, pp. 347-354, 1979.
- [25] R. M. Herman, C. L. Chin, and E. Young, "Nonlinear reflection properties of germanium associated with thermal effects," *Appl. Opt.*, vol. 17, pp. 520-525, Feb. 15, 1978.
- [26] T. A. Wiggins, J. A. Bailey, and A. H. Carieri, "Refractive index changes in germanium due to intense radiation," *Appl. Opt.*, vol. 17, pp. 526-530, Feb. 15, 1978.
- [27] T. A. Wiggins and A. H. Carieri, "Reflectivity changes of Ge due to illumination by means of unequal frequency," *Appl. Opt.*, vol. 18, pp. 1921-1926, June 15, 1979.
- [28] T. A. Wiggins, "Reflectivity changes of Ge and the efficiency of moving diffraction gratings," *Appl. Opt.*, vol. 19, pp. 521-524, Feb. 15, 1980.
- [29] J. H. Lee, J. J. Song, M. A. F. Scarparo, and M. D. Levenson, "Coherent population oscillations and hole burning observed in $\text{Sm}^{2+}:\text{CaF}_2$ using polarization spectroscopy," *Opt. Lett.*, vol. 5, pp. 196-198, May 1980.
- [30] J. H. Lee, "Relaxation time measurements in condensed matter by polarization spectroscopy," Ph.D. dissertation, Univ. Southern California, Los Angeles, 1979.
- [31] R. K. Raj, D. Bloch, J. J. Snyder, G. Camy, and M. Ducloy, "High-frequency optically heterodyned saturation spectroscopy via resonant degenerate four-wave mixing," *Phys. Rev. Lett.*, vol. 44, pp. 1251-1254, May 12, 1980.
- [32] D. Bloch, R. K. Raj, and M. Ducloy, "Doppler-free heterodyne spectroscopy of H_α . Measurement of the $2s_{1/2}$ collisional quenching in gas cell," *Opt. Commun.*, vol. 37, pp. 183-186, May 1, 1981.
- [33] S. Y. Yuen and P. A. Wolff, "Difference-frequency variation of the free-carrier-induced, third-order nonlinear susceptibility in n-InSb," *Appl. Phys. Lett.*, vol. 40, pp. 457-459, Mar. 15, 1982.
- [34] G. Kenney-Wallace, "Subpicosecond tunable spectroscopy: Pulse diagnostics and molecular dynamics in liquids," *Appl. Phys.*, vol. B28, p. 312, June/July 1982.
- [35] G. A. Kenney-Wallace and S. C. Wallace, "Picosecond molecular dynamics in liquids via degenerate four-wave mixing," *IEEE J. Quantum Electron.*, vol. QE-19, pp. 719-723, Apr. 1983.
- [36] M. Golombok and G. A. Kenney-Wallace, "Four wave mixing studies and molecular dynamics simulations," in *Ultrafast Phenomena IV*, Berlin: Springer-Verlag, 1984, pp. 383-386.
- [37] G. A. Kenney-Wallace, "Non-linear optical spectroscopy and molecular dynamics in liquids," in *Applications of Picosecond Spectroscopy to Chemistry*, Reidel, 1984, pp. 139-162.
- [38] M. Golombok, G. A. Kenney-Wallace, and S. C. Wallace, "Pulsed-laser studies of molecular interactions and reorientations of CS_2 in organic liquids via phase conjugation," *J. Phys. Chem.*, to be published.
- [39] R. M. Hochstrasser and C. A. Nyi, "Resonance fluorescence, Raman spectra, and relaxation of single vibronic levels in the condensed phase: Azulene in naphthalene," *J. Chem. Phys.*, vol. 70, pp. 1112-1128, Feb. 1, 1979.
- [40] J. R. Andrews and R. M. Hochstrasser, "Transient grating studies of energy deposition in radiationless processes," *Chem. Phys. Lett.*, vol. 76, pp. 207-212, Dec. 1, 1980.
- [41] —, "Transient grating effects in resonant four-wave-mixing experiments," *Chem. Phys. Lett.*, vol. 76, pp. 213-217, Dec. 1, 1980.
- [42] R. M. Hochstrasser, "Vibrational and electronic relaxation in moderate sized systems," in *Advances in Laser Chemistry*, Berlin: Springer-Verlag, 1978, pp. 98-107.
- [43] J. R. Andrews, R. M. Hochstrasser, and H. P. Trommsdorff, "Vibrational transitions in excited states of molecules using coherent Stokes Raman spectroscopy: Application to ferrocytochrome-C," *Chem. Phys.*, vol. 62, pp. 87-101, 1981.
- [44] D. W. Phillion, D. J. Kuizenga, and A. E. Siegman, "Subnanosecond relaxation time measurements using a transient induced grating method," *Appl. Phys. Lett.*, vol. 27, pp. 85-87, July 15, 1975.
- [45] The presence of such broadening will prevent the determination of τ_2 by simple inverse-linewidth techniques.
- [46] We will work in egs units throughout this work, except when the con-

- ventions are universal, as in watts/cm² for experimental values of intensity.
- [47] J.-L. Oudar and Y. R. Shen, "Nonlinear spectroscopy by multiresonant four-wave mixing," *Phys. Rev. A*, vol. 22, p. 1141, Sept. 1980.
 - [48] P. N. Butcher, *Nonlinear Optical Phenomena*, Bull. 200, Eng. Experiment Station, Ohio State Univ., Columbus, 1964.
 - [49] R. N. Bracewell, *The Fourier Transform and its Applications*, New York: McGraw-Hill, 1965.
 - [50] D. W. Phillion, D. J. Kuizenga, and A. E. Siegman, "Subnanosecond relaxation time measurements using a transient induced grating method," *Appl. Phys. Lett.*, vol. 27, pp. 85-87, July 15, 1975.
 - [51] J. N. Demas, *Excited State Lifetime Measurements*, New York: Academic, 1983.
 - [52] Y. Bard, *Nonlinear Parameter Estimation*, New York: Academic, 1974.
 - [53] N. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed., New York: Wiley, 1981.
 - [54] R. Trebino, "Achromatic N-prism beam expanders: Optimal configurations," *Appl. Opt.*, vol. 24, pp. 1130-1138, Apr. 15, 1985.
 - [55] Th. Förster and G. Hoffman, "Die Viskositätsabhängigkeit der Fluoreszenzquantenausbeuten einiger Farbstoffsysteme," *Z. Phys. Chem. N.G.*, vol. 75, pp. 63-76, 1971.
 - [56] E. P. Ippen and C. V. Shank, "Dynamic spectroscopy and subpicosecond pulse compression," *Appl. Phys. Lett.*, vol. 27, pp. 488-490, Nov. 1, 1975.
 - [57] W. Yu, F. Pellegrino, M. Grant, and R. R. Alfano, "Subnanosecond fluorescence quenching of dye molecules in solution," *J. Chem. Phys.*, vol. 64, pp. 1766-1773, Aug. 15, 1977.
 - [58] V. Sundström and T. Gillbro, "Picosecond dynamics of large-amplitude motions of triphenylmethane molecules in alcohol solution," *Chem. Phys. Lett.*, vol. 110, no. 3, pp. 303-307, 1984.
 - [59] —, "Effects of solvent on TMP photophysics. Transition from no barrier to barrier case, induced by solvent properties," *J. Chem. Phys.*, vol. 81, pp. 3463-3474, Oct. 15, 1984.
 - [60] E. P. Ippen, C. V. Shank, and A. Bergman, "Picosecond recovery dynamics of malachite green," *Chem. Phys. Lett.*, vol. 38, pp. 611-614, Mar. 1976.
 - [61] A. Migus, A. Antonetti, J. Etchepare, D. Hulin, and A. Orszag, "Femtosecond spectroscopy with high-power tunable optical pulses," *J. Opt. Soc. Amer. B*, ser. 2, vol. 2, pp. 584-594, Apr. 1985.
 - [62] D. A. Cremers and M. W. Windsor, "A study of the viscosity-dependent electronic relaxation of some triphenylmethane dyes using picosecond flash photolysis," *Chem. Phys. Lett.*, vol. 71, pp. 27-32, Apr. 1, 1980.
 - [63] J. Salcedo, "The picosecond transient grating technique: Studies on energy transport and photoacoustic effects," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1979.
 - [64] R. Trebino, E. K. Gustafson, and A. E. Siegman, "Fourth-order partial coherence effects in the formation of integrated-intensity gratings with pulsed light sources," submitted to *J. Opt. Soc. Amer. B*.
 - [65] S. Saikan and J. Sei, "Experimental Studies of polarization spectroscopy in dye solutions," *J. Chem. Phys.*, vol. 79, pp. 4146-4153, Nov. 1, 1983.
 - [66] R. C. Desai, M. D. Levenson, and J. A. Barker, "Forced Rayleigh scattering: Thermal and acoustic effects in phase-conjugate wave-front generation," *Phys. Rev. A*, vol. 27, pp. 1968-1976, Apr. 1983.
 - [67] R. Trebino, A. E. Siegman, and C. L. Ladera, "Suppression of thermal gratings in polarization spectroscopy," *J. Opt. Soc. Amer. B*, ser. 2, vol. 1, pp. 549-550, June 1984.
 - [68] P. Nachman, private communication.
 - [69] P. D. Maker, R. W. Terhune, and C. M. Savage, "Intensity-dependent changes in the refractive index of liquids," *Phys. Rev. Lett.*, vol. 12, pp. 507-509, May 4, 1964.
 - [70] D. M. Bloom and G. C. Bjorklund, "Conjugate wave-front generation and image reconstruction by four-wave mixing," *Appl. Phys. Lett.*, vol. 31, pp. 592-594, Nov. 1, 1977.
 - [71] M. A. Duguay and J. W. Hansen, "An ultrafast light gate," *Appl. Phys. Lett.*, vol. 15, pp. 192-194, Sept. 15, 1969.
 - [72] P. W. Smith, J.-P. Hermann, W. J. Tomlinson, and P. J. Maloney, "Optical bistability at a nonlinear interface," *Appl. Phys. Lett.*, vol. 35, pp. 846-848, Dec. 1, 1979.
 - [73] P. A. Madden, "Interaction-induced subpicosecond phenomena in liquids," in *Ultrafast Phenomena IV*, Berlin: Springer-Verlag, 1984, pp. 244-251.
 - [74] Y. Higashigaki, S. L. Whittenburg, and C. H. Wang, "Light scattering studies of orientational fluctuations of CS₂," *J. Chem. Phys.*, vol. 69, pp. 3297-3301, Oct. 1, 1978.
 - [75] T. I. Cox, M. R. Battaglia, and P. A. Madden, "Properties of liquid CS₂ from the allowed light scattering spectra," *Molec. Phys.*, vol. 38, no. 5, pp. 1539-1554, 1979.
 - [76] J.-M. Talbot and C. L. Tang, "Femtosecond interferometry for nonlinear optics," *Appl. Phys. Lett.*, vol. 40, pp. 765-767, May 1, 1982.
 - [77] B. I. Greene and R. C. Farrow, "Femtosecond transient birefringence in CS₂," in *Picosecond Phenomena III*, Berlin: Springer-Verlag, 1982, pp. 209-211.
 - [78] B. J. Berne and R. Pecora, *Dynamic Light Scattering*, New York: Wiley, 1976.
 - [79] R. W. Hellwarth, "Third-order optical susceptibilities of liquids and solids," *Progr. Quantum Electron.*, vol. 5, pp. 1-68, 1977.
 - [80] C. V. Shank, private communication.
 - [81] G. Steel and S. C. Rand, "Ultrasharp nonlinear optical resonances in solids," *Phys. Rev. Lett.*, to be published.



Rick Trebino was born in Boston, MA, on January 18, 1954. He received the B.A. degree magna cum laude in physics from Harvard University, Cambridge, MA, in 1977, and the M.S. and Ph.D. degrees in applied physics from Stanford University, Stanford, CA. His dissertation research consisted of the development of a novel technique for the measurement of femtosecond events using frequency-domain nondegenerate four-wave mixing.

Currently, he is a Physical Sciences Research Associate with the Applied Physics Department, Stanford University, where he is continuing his research on femtosecond processes, specializing in internal conversion in triphenyl methane dyes and orientational relaxation in optical-Kerr materials. He is also studying higher order partial-coherence effects in nonlinear optical processes, achromatic multiprism beam expanders, and pulsed-dye-laser engineering. In addition to his position at Stanford, he is a consultant in medical physics for numerous firms in the San Francisco area and, at the same time, is the President of the Stanford Laser Consulting Group. He holds two patents for optical devices and has received fellowships from IBM and ARCS.

Charles E. Barker, photograph and biography not available at the time of publication.

Anthony E. Siegman (S'54-M'57-F'66), for a photograph and biography, see this issue, p. 1403.

Fourth-order partial-coherence effects in the formation of integrated-intensity gratings with pulsed light sources

Rick Trebino, Eric K. Gustafson, and A. E. Siegman

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305

Received December 20, 1985; accepted May 5, 1986

We performed theoretical calculations of the relative diffraction efficiency of partially coherent light-induced integrated-intensity gratings using pulsed sources, paying particular attention to thermal gratings. We provided a simple intuitive picture of the phenomenon and then calculated exact expressions that, unlike instantaneous-intensity-grating results, necessarily require the use of fourth-order coherence functions. Assuming several radiation models, we evaluated these expressions and found that the results proved to be insensitive to the specific radiation model assumed. The application of these results to a previously performed pulsed-laser experiment yielded a better fit to the data than an expression involving only second-order coherence, which is valid only in the cw limit. We included the effects of grating decay and, in addition, compared the use of integrated-intensity gratings for ultrashort-pulse-length measurement with standard techniques. Finally, we calculated expressions for the relative diffraction efficiency of integrated-intensity gratings created with excitation beams from two separate and independent sources of different frequency, and we report an experiment whose results were found to agree with this theory.

1. INTRODUCTION

Integrated-intensity gratings, induced in a material by interfering light beams, arise in many nonlinear-optical experimental techniques. In some degenerate four-wave mixing materials, diffraction from integrated-intensity gratings yields efficient phase-conjugate reflection.^{1,2} In short-pulse transient-grating experiments, the creation of integrated-intensity gratings permits the measurement of diffusivity constants and also the study of acoustic waves created in the process.³⁻⁵ In addition, the large magnitude of some integrated-intensity-grating effects has led to their use in schemes for the measurement of ultrashort pulse lengths⁶ and coherence times.^{7,8} On the negative side, integrated-intensity gratings in the form of thermal gratings often prove to be a nuisance in laser techniques intended to measure population, but not thermal, effects.⁹ And in short-pulse excite-probe experiments, the "coherence spike" often originates from a thermal grating or other integrated-intensity grating.¹⁰

In all light-induced-grating interactions, the partial-coherence properties of the excitation beams play an important role. We have known from the time of Michelson that the ability to form integrated-intensity fringes requires good temporal coherence between the excitation beams.¹¹ A single light source, providing both beams with a relative delay of less than each beam's coherence time, is generally required for intensity fringes to be obtained. The use of a large delay generally results in weak fringes, and worse, two separate sources produce no fringes at all.

When integrated-intensity fringes arise in *pulsed-laser*, induced-grating experiments, however, the situation is somewhat different from the well-known example mentioned above. First, rather than using radiation of infinite duration, these experiments employ a series of pulses, each pulse being a finite number of coherence times in length. Second, rather than the actual fringes, what is observed is a

beam diffracted by the fringes. And by measuring only the *intensity* of the diffracted beam, we lose the information on its *phase*, which consequently deprives us of information on the induced-grating phase.

Why are these differences important? The loss of phase information is important because, by measuring only the amplitude of the grating fringes, we measure an inherently positive quantity. And by using excitation pulses a finite number of coherence times in duration, the total washout of the intensity fringes that can take place in cw experiments does not occur. On each pulse, then, we measure an inherently positive diffraction efficiency. That positive quantity is then averaged over many pulses to yield a potentially significant signal. Thus pulsed-laser, integrated-intensity-grating experiments can yield large diffraction efficiencies, while cw experiments, or those that are sensitive to the instantaneous phase of the fringes, will measure very little fringe strength.

In this paper we give an intuitive picture of integrated-intensity-grating formation with partially coherent light pulses. We derive theoretical expressions for the diffraction efficiencies of integrated-intensity gratings, and we show that these expressions necessarily involve fourth-order coherence functions. This mathematical theory differs from the second-order coherence-function theory of the Michelson interferometer and related fringe-producing devices¹¹ because induced-grating experiments measure the *squared* amplitude of the fringes: the diffracted radiation *field* is proportional to the complex grating amplitude, and the diffracted *intensity* is what is measured. Because the complex grating amplitude is proportional to the product of the two excitation-beam envelopes, the product of *four* radiation fields will occur in any expression for the observed diffracted intensity. Any theoretical treatment of induced gratings using partially coherent light pulses will then necessarily involve *fourth-order* coherence functions.

This fourth-order theory reduces in the cw limit to the

well-known second-order theory; the deviations from the second-order theory will be proportional to the ratio of the coherence time, τ_c , to the pulse length, τ_p , of the excitation radiation. Only in the past few years have light sources achieved Fourier-transform-limited quality, for which $\tau_c/\tau_p \sim 1$, so that the above deviations from the second-order theory have only recently become observable. And, in fact, several recently performed pulsed experiments exhibit deviations from the second-order theory. A variable-delay, integrated-intensity-grating experiment performed by Eichler *et al.*⁷ exhibited greater diffraction for large delays than were predicted by the second-order theory. In addition, various thermal-grating experiments employing excitation beams from two separate lasers—for which the second-order theory predicts no diffraction at all—observe such strong diffraction that other (instantaneous-intensity) effects have been obscured.^{9,12} The fourth-order theory developed herein will accurately explain these experiments.

We will consider the special case of *thermal* gratings because typical thermal-grating decay times are generally quite long compared with *Q*-switched- and mode-locked-laser pulses, so that thermal gratings induced by such pulses are in general modeled accurately as integrated-intensity gratings. Other types of induced gratings will also behave as integrated-intensity gratings, and the quantities used here can be relabeled easily to describe them.

We derive exact expressions for the relative magnitude of the (integrated-intensity) thermal-grating diffraction efficiency for several radiation models, specifically, amplitude-stabilized and thermal radiation with Lorentzian and Gaussian line shapes. For thermal radiation, these expressions reduce to particularly useful and intuitive results when the radiation coherence time is much less than the pulse length. In this regime, our expressions contain two terms, one equal to the squared magnitude of the second-order coherence function and the other proportional to the pulse envelope autocorrelation function. Applying these results to a previously performed experiment,⁷ we obtain a better fit than that obtained with previous theories.

In addition, we discuss the merits of thermal-grating techniques for the measurement of laser pulse lengths and coherence times and compare these techniques with standard methods. Finally, we calculate the relative diffraction efficiency of an integrated-intensity grating using two independent excitation sources, and we report an experiment whose results agree with the predictions of this theory.

Previous theoretical work on thermal and other integrated-intensity gratings induced with pulsed partially coherent radiation include that of Vardeny and Tauc,¹⁰ who obtain some of the general expressions contained here but do not consider specific radiation models. Eichler *et al.*⁷ do consider specific models for the input radiation but employ a second-order theory, obtaining only coherence effects and not pulse-length effects. Idiatulin and Teryaev,⁶ on the other hand, obtain the pulse autocorrelation term but not the coherence term. Many authors have treated the so-called "coherent-coupling spike" of ultrashort-pulse, excite-probe experiments¹³⁻¹⁶ and some have included integrated-intensity effects and fourth-order effects, but none have considered specific radiation models. Grossman and Shemwell¹⁷ calculated the effects of poor coherence on phase conjugation by degenerate four-wave mixing. Finally, fourth- and

higher-order coherence functions arise in a number of phenomena; numerous calculations involving them exist in the literature.^{11,18,19}

2. PRELIMINARY THEORY: AN INTUITIVE PICTURE AND A REVIEW OF INTEGRATED-INTENSITY-GRATING THEORY

A simple argument shows how, despite poor coherence between excitation beams, a pulsed-laser-induced, integrated-intensity grating will diffract light and, in addition, approximately how much light will be diffracted. Suppose the excitation pulses are N coherence times long. The phase of the intensity-fringe pattern will change randomly (and, we assume for simplicity, discretely) on the scale of a coherence time. After N coherence times, the integrated-intensity-grating complex amplitude, A , will be the sum of the individually contributed gratings from each coherence-time period:

$$A = \sum_{m=1}^N \exp(i\phi_m), \quad (1)$$

where ϕ_m is the phase of the fringe pattern during the m th period. The expected diffraction efficiency, $\langle \eta \rangle$, is proportional to the ensemble expectation of $|A|^2$ and will be

$$\langle \eta \rangle \propto \left\langle \left| \sum_{m=1}^N \exp(i\phi_m) \right|^2 \right\rangle. \quad (2)$$

For randomly distributed individual phases ϕ_m , the expectation value takes on the value N . Thus, the diffraction efficiency will not be zero. For comparison, suppose that the coherence between the beams is perfect. Then, throughout the pulse (continue to break the pulse down into N increments), all contributions to the grating will be in phase, i.e., all ϕ_m are equal. As a result, in this case, $\langle \eta \rangle \propto N^2$. Comparing the diffraction efficiencies of these two cases indicates that poor coherence results in a lessened diffraction efficiency by a factor of N , the ratio of the pulse length to the coherence time. In the cw limit, $N \rightarrow \infty$, and this reduction is infinite, but in pulsed-laser experiments N will be a finite number and a considerable diffraction efficiency will remain.²⁰

This argument shows that when the excitation sources of an integrated-intensity grating are independent, the diffraction efficiency of the grating will be reduced by a factor of the order of τ_c/τ_p , where τ_c is the coherence time and τ_p is the pulse length. The argument applies to experiments employing one light source in which the relative delay between the two excitation beams derived from this source is much greater than the source coherence time. It also applies to experiments employing two independent light sources of the same frequency. Thus, despite the minimal coherence between the two excitation beams in these two types of experiment, a finite diffraction efficiency will exist.

To formalize the above argument and to allow for specific radiation models, we must take into account the actual time dependences of the excitation fields. Consequently, we consider two quasi-monochromatic plane waves of the same frequency (ω) and polarization (\hat{e})

$$\mathcal{E}_1(\mathbf{r}, t) = \text{Re } E_1(t) \exp i(\omega t - \mathbf{k}_1 \cdot \mathbf{r}) \hat{e},$$

$$\mathcal{E}_2(\mathbf{r}, t) = \text{Re } E_2(t) \exp i(\omega t - \mathbf{k}_2 \cdot \mathbf{r}) \hat{e},$$

where $E_1(t)$ and $E_2(t)$ are slowly varying complex amplitudes. Suppose that these two beams of light simultaneously illuminate an absorbing material. In cgs units, the intensity in the material will be

$$I(\mathbf{r}, t) = I_1(t) + I_2(t) + \text{Re} \frac{c}{8\pi} (\epsilon/\mu)^{1/2} E_1(t) E_2^*(t) \times \exp(-i\mathbf{k}_{gr} \cdot \mathbf{r}), \quad (3)$$

where $I_1(t)$ and $I_2(t)$ are the individual beam intensities, $\mathbf{k}_{gr} \equiv \mathbf{k}_1 - \mathbf{k}_2$ is the grating k vector, ϵ is the permittivity of the medium, μ is the magnetic permeability, and c is the speed of the light in vacuum.

We will now consider a special case of integrated-intensity gratings: the phase grating resulting from the combination of a nonzero dn/dT , where n is the material refractive index and T is the temperature, and the spatially sinusoidal deposition of heat in an absorbing material. Thus, we require the energy density deposited in the absorbing material in a time interval, τ_p (i.e., the pulse length),

$$W(\mathbf{r}) = \alpha \int_{-\tau_p/2}^{\tau_p/2} I(\mathbf{r}, t) dt, \quad (4)$$

where α is the material absorption coefficient, and we will assume that only a small fraction of the beam intensity is absorbed by the material so that the intensity remains independent of position in directions perpendicular to \mathbf{k}_{gr} . If Φ is the fraction of the deposited energy that becomes heat, then the resulting temperature distribution in the material will be

$$T(\mathbf{r}) = T_0 + \left(\frac{\Phi}{\rho c_v} \right) W(\mathbf{r}), \quad (5)$$

where T_0 is the ambient temperature, ρ is the material density, and c_v is the specific heat at constant volume. We assume that no heat-dissipative effects take place on the time scale, τ_p . Substituting Eqs. (3) and (4) into Eq. (5), we obtain both a uniform increase in the material temperature, ΔT_0 , and a spatially modulated term, ΔT_{gr} ,

$$T(\mathbf{r}) = T_0 + \Delta T_0 + \text{Re} \Delta T_{gr} \exp(-i\mathbf{k}_{gr} \cdot \mathbf{r}), \quad (6)$$

where

$$\Delta T_0 = \frac{\Phi \alpha}{\rho c_v} \int_{-\tau_p/2}^{\tau_p/2} [I_1(t) + I_2(t)] dt \quad (7)$$

and

$$\Delta T_{gr} = \frac{\Phi \alpha}{\rho c_v} 2 \frac{c}{8\pi} (\epsilon/\mu)^{1/2} \int_{-\tau_p/2}^{\tau_p/2} E_1(t) E_2^*(t) dt. \quad (8)$$

The thermal grating arises directly from the spatially modulated term.

Most materials exhibit a temperature-dependent refractive index. Consequently the spatial temperature modulation will become a spatial refractive-index modulation, or phase grating, with amplitude Δn_{gr} :

$$\Delta n_{gr} = \left[\left(\frac{dn}{dT} \right) \frac{\Phi \alpha}{\rho c_v} 2 \frac{c}{8\pi} (\epsilon/\mu)^{1/2} \int_{-\tau_p/2}^{\tau_p/2} E_1(t) E_2^*(t) dt \right], \quad (9)$$

where dn/dT is the derivative of refractive index with respect to temperature, evaluated at $T_0 + \Delta T_0$. The diffraction efficiency, η , of this phase grating will be

$$\eta = K^2 \left| \int_{-\tau_p/2}^{\tau_p/2} E_1(t) E_2^*(t) dt \right|^2, \quad (10)$$

where

$$K = \left(\frac{dn}{dT} \right) \frac{\Phi \alpha}{\rho c_v} 2 \frac{c}{8\pi} (\epsilon/\mu)^{1/2} k_{pr} L. \quad (11)$$

k_{pr} is the probe-beam k vector and L is the sample length.¹² We have assumed the low-diffraction-efficiency limit. If we now explicitly assume *monochromatic* beams, whose field amplitudes, $E_i(t)$, will be constant, Eq. (10) becomes

$$\eta = K^2 |E_1|^2 |E_2|^2 \tau_p^2. \quad (12)$$

Use of Eq. (12) for *relative* numbers yields interesting and useful information.²¹ *Absolute* estimates obtained by using Eq. (12), however, usually give results that are several orders of magnitude larger than experimental values.

Several factors contribute to lowering the efficiency of a thermal grating from this theoretical value. First, the deposition of heat into the sample may require a significant amount of time, so that if relatively short pulses are used and probing occurs simultaneously with or just after excitation, the temperature-modulation buildup will not be complete until after probing occurs. Second, in general, the temperature grating must generate a *density* grating to cause the refractive-index modulation. This process also takes time and will contribute to the observation of a much weaker grating than that predicted by Eq. (12) in short-pulse experiments.²² Third, spatial variations in the excitation-beam intensities can also result in significantly less efficiency than would be expected from a spatially-flat intensity distribution.²³ Finally, temporal variations in the excitation-beam fields in both amplitude and phase will also act to change the efficiency. Of particular importance are temporal variations in the phases of these beams, which can result in a variation of the induced-grating phase and a washing out of the induced grating. These phase variations are similar to those that result in the loss of fringe visibility in the Michelson interferometer.

The inclusion of all the above effects is beyond the scope of this work, and as a result we will not attempt to predict *absolute* diffraction efficiencies but only *relative* efficiencies. We have, however, already included the effects of temporal variations in the excitation-beam fields in Eq. (10), which we will rewrite here as

$$\eta = K^2 \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} E_1(t_1) E_2^*(t_1) E_1^*(t_2) E_2(t_2) dt_1 dt_2. \quad (13)$$

Rewriting this equation in terms of dimensionless field quantities, $u_i(t)$, we have

$$\eta = K^2 |E_1|^2 |E_2|^2 \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} u_1(t_1) u_2^*(t_1) u_1^*(t_2) u_2(t_2) dt_1 dt_2, \quad (14)$$

in which $E_i(t) = E_i u_i(t)$, where E_i is a constant field magnitude and $u_i(t)$ contains the time dependence of the excitation-beam fields, i.e., phase and amplitude fluctuations. The expectation value of the modulus of $u_i(t)$ is assumed to be unity.

We will compare results obtained by using Eq. (14) with those of Eq. (12), which is an ideal case involving monochro-

matic radiation (although averaged for a time, τ_p , only). In all comparisons, the several effects mentioned above, which affect the value of K , will cancel when we normalize by the monochromatic-beam diffraction efficiency, $K^2|E_1|^2|E_2|^2\tau_p^2$. So define the normalized thermal-grating diffraction efficiency, $\tilde{\eta}$:

$$\tilde{\eta} = \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} u_1(t_1)u_2^*(t_1)u_1^*(t_2)u_2(t_2)dt_1dt_2. \quad (15)$$

We must recognize that the thermal-grating diffraction efficiency is a statistical quantity, that is, variations in the diffraction efficiency will occur on a shot-to-shot basis owing to variations in the excitation-beam temporal waveforms. Most experiments measure averages over many shots; it is therefore necessary to consider the ensemble expectation of the normalized diffraction efficiency:

$$\langle \tilde{\eta} \rangle = \left\langle \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} u_1(t_1)u_2(t_2)u_1^*(t_2)u_2^*(t_1)dt_1dt_2 \right\rangle, \quad (16)$$

where the brackets denote the ensemble expectation operator and we have permuted the factors within the brackets. The time-integration and expectation operators commute, so we have

$$\langle \tilde{\eta} \rangle = \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} \langle u_1(t_1)u_2(t_2)u_1^*(t_2)u_2^*(t_1) \rangle dt_1dt_2, \quad (17)$$

which is similar to results derived by others.^{6,10,13} The derivation of this result does not depend on the specific properties of the statistics of the radiation field. The product of four fields has arisen because the measured quantity is the square of the fringe intensity-pattern amplitude.

3. THERMAL GRATINGS INDUCED BY PARTIALLY COHERENT LIGHT FROM A SINGLE LASER

Suppose that both excitation beams emanate from the same laser or light source but that one beam traverses a greater distance than the other before reaching the sample material (see Fig. 1). Assume further that the delay between the two beams, τ_d , is much less than the pulse length, τ_p , so that excitation-pulse overlap in time is good. We have, therefore, $u_2(t) = u_1(t + \tau_d)$, so that the integrand in Eq. (17) becomes $\langle u_1(t_1)u_1(t_2 + \tau_d)u_1^*(t_2)u_1^*(t_1 + \tau_d) \rangle$, which is the fourth-order coherence function, $\Gamma^{(4)}(t_1, t_2 + \tau_d; t_2, t_1 + \tau_d)$ for the field $u_1(t)$. Equation (17) now becomes

$$\langle \tilde{\eta} \rangle = \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} \Gamma^{(4)}(t_1, t_2 + \tau_d; t_2, t_1 + \tau_d) dt_1dt_2. \quad (18)$$

Equation (18) is the ratio of two quantities: the first is the expected efficiency of a thermal grating probed after its formation, when both excitation pulses emanate from the same partially coherent source with relative delay, τ_d , between them, and the second is the diffraction efficiency of a similar grating formed with monochromatic excitation beams. To proceed further, we must assume a statistical model for the partially coherent excitation radiation. It is not sufficient to know merely the frequency spectrum of the excitation radiation, because the recursion relations relating the fourth-order coherence function to second-order coherence functions depend on the radiation model used.

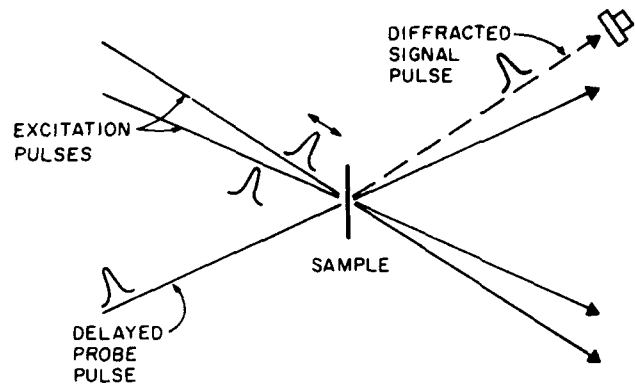


Fig. 1. Experimental arrangement for the study of integrated-intensity gratings. Two excitation pulses experience a variable delay with respect to each other and excite the grating. The probe pulse, further delayed, probes the induced grating, and a detector detects the diffracted beam.

We will use various models for the excitation-beam statistics, including an amplitude-stabilized quasi-monochromatic source²⁴ (which is often used to approximate a single-mode laser) with a Lorentzian frequency spectrum and also a thermal light source¹¹ (which approximates multimode laser sources) with Lorentzian and Gaussian spectra. All these theoretical radiation fields represent ergodic processes, while the pulsed nature of the experiment, which could otherwise destroy ergodicity, is included by integrating for the time τ_p . Later, pulse-shape effects will be included by the inclusion of a deterministic pulse-shape function, $A(t)$, with the random process, $u(t)$, still retaining the convenient property of ergodicity. In general, however, laser light does not represent an ergodic statistical process. And, in particular, the above models do not accurately describe certain types of laser radiation, e.g., mode-locked-laser pulses.¹¹ It should be noted here, though, that each of the above models yields approximately the same quantitative results, which agree well with both intuition and experiment and are therefore adequate for many purposes.

The recursion relation for the fourth-order coherence function of an amplitude-stabilized quasi-monochromatic field is²⁴

$$\Gamma^{(4)}(t_1, t_2; t_3, t_4) = \Gamma^{(2)}(t_1 - t_3)\Gamma^{(2)}(t_2 - t_4) \times \left[\frac{\Gamma^{(2)}(t_1 - t_4)\Gamma^{(2)}(t_2 - t_3)}{\Gamma^{(2)}(t_1 - t_2)\Gamma^{(2)}(t_3 - t_4)} \right], \quad (19)$$

and, for thermal radiation, the recursion relation is¹¹

$$\Gamma^{(4)}(t_1, t_2; t_3, t_4) = \Gamma^{(2)}(t_1 - t_3)\Gamma^{(2)}(t_2 - t_4) + \Gamma^{(2)}(t_1 - t_4)\Gamma^{(2)}(t_2 - t_3). \quad (20)$$

The required integration of these quantities for radiation with Lorentzian and Gaussian line shapes is not difficult, and exact results for various cases are given in Appendix A. The results for all models considered are similar. For the purposes of discussion, we consider thermal radiation with a Lorentzian line shape (a reasonable approximation of a multimode pulsed dye laser). The expected normalized diffraction efficiency to first order in τ_d/τ_p is

$$\langle \tilde{\eta} \rangle \approx \frac{\tau_c}{\tau_p} + \exp(-2|\tau_d/\tau_c|) = \frac{\tau_c}{\tau_p} + |\Gamma^{(2)}(\tau_d)|^2. \quad (21)$$

The exponential (the second-order coherence function) in expression (21) is the well-known "second-order" result,⁷ while the other term corresponds to the background, which results from the incomplete washout of the grating in the finite number of coherence times in the pulse length, despite the relative incoherence of the beams. A few limiting cases are of interest:

(1) When the delay, τ_d , is zero, $\langle \tilde{\eta} \rangle \approx 1$, as expected: when the delay is zero, the relative phase of the two excitation beams does not change, and no grating washout occurs; this situation is equivalent to that of monochromatic excitation beams.

(2) When the pulse length, τ_p , approaches infinity, our result approaches the well-known result for the cw limit: $\langle \tilde{\eta} \rangle = \exp(-2|\tau_d/\tau_c|) = |\Gamma^{(2)}(\tau_d)|^2$. In this case, grating washout will occur for large delays ($\tau_d \gg \tau_c$) but not for small delays ($\tau_d \ll \tau_c$), as expected.

(3) When the coherence time, τ_c , is close to zero, the intensity pattern may exist for a finite time, but its phase dances back and forth extremely fast, again washing out the grating: $\langle \tilde{\eta} \rangle = 0$ for all nonzero values of delay.

(4) The most interesting limit is the case $\tau_d \gg \tau_c$, but with τ_p finite. We then have $\langle \tilde{\eta} \rangle \approx \tau_c/\tau_p$. This means that, despite using a delay much longer than the coherence time, a nonzero expected diffraction efficiency results. This occurs because the finite pulse length allows only τ_p/τ_c coherence times for washout to occur, and as a result, washout will not be complete. This limit illustrates the primary difference between pulsed and cw experiments, and it shows that thermal gratings can exist in experiments in which excitation beams are drawn from *separate* lasers (equivalent to the $\tau_d \rightarrow \infty$ limit), provided that their average frequencies are equal.

When the delay between pulses approaches the laser pulse length, poor pulse overlap begins to occur and pulse-shape effects must be included in the analysis. When this happens, the τ_c/τ_p -background level will decrease, and when $\tau_d \gg \tau_p$, the background will approach zero because the excitation beams cease to overlap at all. It is easy to see that, if we write the field as $E(t) = FA(t)u(t)$, in which $u(t)$ is the statistical factor that includes the phase of the field and $A(t)$ is a normalized, real, deterministic amplitude function, Eq (18) becomes

$$\langle \tilde{\eta} \rangle = \frac{1}{\tau_p^2} \int_{-\tau_p}^{\tau_p} \int_{-\tau_p}^{\tau_p} A(t_1)A(t_2 + \tau_d)A(t_1 + \tau_d)A(t_2) \times \Gamma^{(4)}(t_1, t_2 + \tau_d; t_1 + \tau_d, t_2) dt_1 dt_2. \quad (22)$$

This expression is also easily evaluated for the various radiation models of interest, and Appendix B contains exact expressions for thermal radiation with Gaussian and hyperbolic-secant-squared pulse shapes. We can, however, derive a simple and intuitive expression, analogous to relation (21), when $\tau_c \ll \tau_p$. If we assume thermal radiation, we obtain

$$\langle \tilde{\eta} \rangle \approx \frac{\tau_c}{\tau_p} \int_{-\tau_p}^{\tau_p} I(t)I(t + \tau_d)dt + |\Gamma^{(2)}(\tau_d)|^2, \quad (23)$$

independent of the precise shapes of the pulse and line and where $I(t)$ is the normalized intensity: $I(t) \equiv A^2(t)$ defined

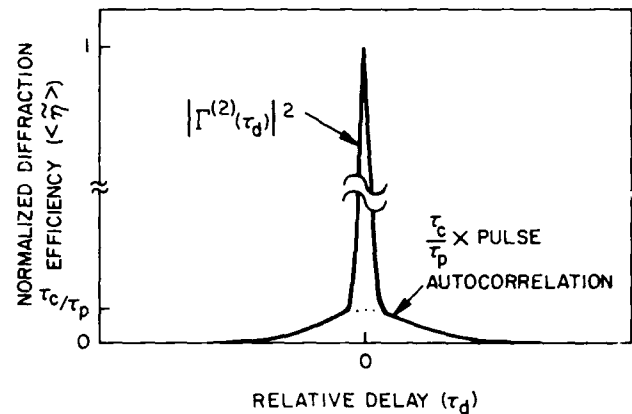


Fig. 2. Plot of theoretical diffraction efficiency versus delay between excitation beams for the case of $\tau_c \ll \tau_p$.

so that when $\tau_d = 0$, the integral in expression (23) is unity. As before, the second term is the well-known second-order result,⁷ and the first term is the background, now seen to be the pulse autocorrelation function. The magnitude of the first term is still τ_c/τ_p smaller than the second, but now only when $\tau_d = 0$; when $|\tau_d|$ increases to values on the order of τ_p , the first term's strength decreases (see Fig. 2).

4. COMPARISON WITH EXPERIMENT

These expressions describe a grating experiment performed by Eichler *et al.*,⁷ in which picosecond pulses from a frequency-doubled Nd:YAG laser were split and interfered to form an integrated-intensity grating (probably thermal but possibly having a long-lived population component, also) in a solution of Rhodamine 6G dissolved in ethanol. Eichler *et al.* fitted their data to a second-order theory, and their best fit is shown as the dashed line in Fig. 3. We have fitted their data to a fourth-order theory, shown by the solid line, resulting from a hyperbolic-secant-squared pulse shape and thermal Lorentzian line and using expression (23):

$$\langle \tilde{\eta} \rangle \approx \frac{\zeta \tau_c}{\tau_p} \left[\frac{\zeta \tau_d / \tau_p \cosh(\zeta \tau_d / \tau_p) - \sinh(\zeta \tau_d / \tau_p)}{\sinh^3(\zeta \tau_d / \tau_p)} \right] + \exp(-2|\tau_d/\tau_c|), \quad (24)$$

where $\zeta = 1.7627$. The fourth-order theory yields a much better fit. We derive the pulse length from the width of the "background" to be 26.5 ± 0.5 psec, a bit higher than the value measured by Eichler *et al.* with a streak camera: 22 ± 4 psec. The smaller error bars accompanying our value reflect the multishot averaging in the grating experiment (compared to the single-shot nature of the streak camera measurement). We also derive a coherence time, τ_c , of 8.5 ± 0.5 psec from the width of the central spike, also larger than the experimental value measured by Eichler *et al.* of 2.7 psec, and obtained from the expression $\tau_c = 1/\pi\delta\nu$, where $\delta\nu = 1.2 \times 10^{11}$ Hz. This discrepancy is probably due to the fact that Eichler *et al.* measured $\delta\nu$ by using different pump powers and alignments than were used in this experiment. Significantly, the theory fits the data reasonably well, despite the fact that the ratio of the magnitudes of the two terms is determined by these two times and cannot be independently curve fitted.

It should be noted that, in using expression (23), we have

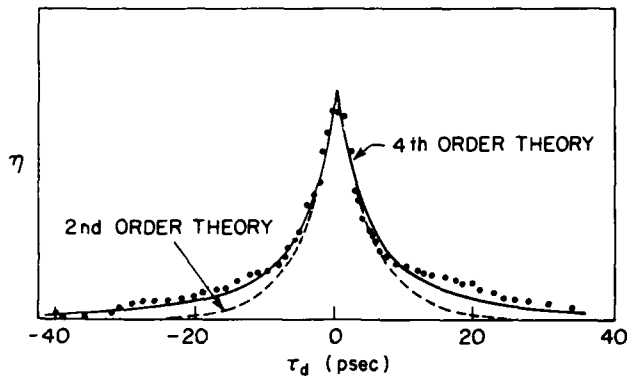


Fig. 3. Experimental integrated-intensity-grating diffraction efficiency versus relative delay in an experiment of Eichler *et al.*⁷ The dashed curve represents the best fit to a (cw-limit) second-order theory, while the solid curve represents the best fit to an approximation to the fourth-order theory developed here. Note that the fourth-order theory yields a much better fit in the wings of the data, where only incomplete washout of the grating occurs, and poor pulse-overlap limits the diffraction efficiency.

employed a theory that assumes that $\tau_c \ll \tau_p$ to fit data in which these quantities apparently differ by only a factor of approximately 3. The error introduced by this assumption will be of the order of τ_c^2/τ_p^2 , and, in general, the coefficient multiplying this quantity is less than unity (see Appendix B). The maximum error anticipated here is a few percent. Significantly, the use of exact results for radiation with a less appropriate (Gaussian) pulse shape [Eq. (B5)] fit the data equally well. The error introduced by the use of a thermal-radiation model to describe the mode-locked pulses is probably larger.

5. PULSE-LENGTH MEASUREMENT USING THERMAL GRATINGS

Eichler *et al.* performed their experiment to demonstrate that integrated-intensity gratings can be used to measure radiation coherence times. We have observed that their experiment also provides an autocorrelation measurement for the pulse length. Because thermal gratings can be quite efficient and because automatically phase-matched geometries (such as the polarization spectroscopy geometry used by Song *et al.*²⁵ and some background-free geometries employing counterpropagating excitation and probe beams²⁶) exist for the performance of thermal-grating experiments, such a technique may be preferred over current techniques, such as second-harmonic generation (SHG).

We can compare this thermal-grating pulse-autocorrelation method with SHG methods by noting that the expected normalized SHG energy efficiency $\langle \tilde{\eta}_{\text{SHG}} \rangle$ can be written as

$$\langle \tilde{\eta}_{\text{SHG}} \rangle = \int_{-\infty}^{\infty} \langle |A(t)A(t + \tau_d)u(t)u(t + \tau_d)|^2 \rangle dt, \quad (25)$$

where as before $A(t)$ represents a smooth deterministic pulse-shape envelope and $u(t)$ contains the phase and amplitude fluctuations. Rewriting this expression, we have

$$\langle \tilde{\eta}_{\text{SHG}} \rangle = \int_{-\infty}^{\infty} A^2(t)A^2(t + \tau_d)\Gamma^{(4)}(t, t + \tau_d; t, t + \tau_d)dt. \quad (26)$$

This expression is analogous to Eq. (22), which gives the

normalized thermal-grating diffraction efficiency if probing occurs after its formation.

We would like to simplify these expressions further. Observing that, for Fourier-transform-limited pulses, $\tau_c \approx \tau_p$, and partial-coherence effects will be absent (both methods yield the pulse autocorrelation without central spikes). Thus we must consider the $\tau_c \ll \tau_p$ regime. To do so, a radiation model must be chosen. We will again choose a thermal model, for simplicity but also because, when $\tau_c \ll \tau_p$, good mode locking has evidently not occurred and considerable randomness in the modes' phases must exist. Copying the result for thermal gratings, expression (23), and using the recursion relation for $\Gamma^{(4)}$, Eq. (20), we obtain

$$\langle \tilde{\eta}_{\text{TG}} \rangle \approx \frac{\tau_c}{\tau_p} \int_{-\infty}^{\infty} I(t)I(t + \tau_d)dt + |\Gamma^{(2)}(\tau_d)|^2 \quad (27)$$

and

$$\langle \tilde{\eta}_{\text{SHG}} \rangle \approx \int_{-\infty}^{\infty} I(t)I(t + \tau_d)dt + |\Gamma^{(2)}(\tau_d)|^2, \quad (28)$$

when $\tau_c \ll \tau_p$; TG stands for thermal grating. Thus the autocorrelation function is weaker in thermal-grating methods by a factor of τ_c/τ_p . For pulses that are far from Fourier-transform limited, the autocorrelation may be difficult to see next to the central spike. That this background attenuation occurs is no surprise: thermal-grating methods are sensitive to phase variations in the beams, and the resultant fringe washout reduces diffraction efficiency. It must also be admitted that sensitivity to phase is in general not desirable in a pulse-amplitude measurement technique. Thus SHG or other methods, such as two-photon fluorescence, that measure only the pulse amplitude yield more desirable information, while thermal-grating methods are less expensive and are easier to perform.

On the other hand, because of the reduced autocorrelation background due to this phase sensitivity, thermal-grating techniques may provide an excellent measure of radiation coherence times, as suggested by Eichler *et al.* It must be remembered, however, that one measures a fourth-order coherence function, not a second-order coherence function, and unless the experimenter knows the appropriate model, i.e., recursion relation, for his radiation, it may not be possible to extract $\Gamma^{(2)}(t)$ from the measurement.

6. INCLUSION OF GRATING-RELAXATION EFFECTS

If the grating decay time is of the order of the laser pulse length, a new theory incorporating this decay, the laser pulse shape, and perhaps partial coherence effects must be developed. This is easily done,^{10,12} and Eq. (17) becomes

$$\begin{aligned} \langle \tilde{\eta}(t) \rangle = & \int_{-\infty}^t \int_{-\infty}^t A(t_1)A(t_2 + \tau_d)A(t_2)A(t_1 + \tau_d) \\ & \times \Gamma^{(4)}(t_1, t_2 + \tau_d; t_2, t_1 + \tau_d)h(t - t_1)h(t - t_2)dt_1dt_2, \end{aligned} \quad (29)$$

where $h(t)$ is the grating decay function. The diffraction efficiency is now necessarily time dependent because, using pulses, the grating will eventually decay away completely. Usually the grating decay function will be exponential: $h(t)$

$= \exp(-t/\tau_{th})\theta(t)$, where τ_{th} is the decay time scale and $\theta(t)$ is the unit step function. We assume thermal radiation and again work in the regime $\tau_c \ll \tau_p$. Also, we assume square pulses, $A(t) = \text{rect}(t/\tau_p)$, for simplicity, and calculate the diffraction efficiency for points in time after the two pulses have passed through the sample material. For delays that yield at least some pulse overlap,

$$\langle \tilde{\eta}(t) \rangle = \exp\left(-\frac{2t + \tau_d}{\tau_{th}}\right) \left[\tau_c \tau_{th} \sinh\left(\frac{\tau_p - |\tau_d|}{\tau_{th}}\right) + 4\tau_{th}^2 |\Gamma^{(2)}(\tau_d)|^2 \sinh^2\left(\frac{\tau_p - |\tau_d|}{2\tau_{th}}\right) \right] \quad (30)$$

We can more easily compare Eq. (30) with previous results if we renormalize so that the coefficient of $|\Gamma^{(2)}(\tau_d)|^2$ is one and also assume that $|\tau_d| \ll \tau_p$:

$$\langle \tilde{\eta}(t) \rangle \propto \frac{\tau_c}{4\tau_{th}} \frac{\sinh(\tau_p/\tau_{th})}{\sinh^2(\tau_p/2\tau_{th})} + |\Gamma^{(2)}(\tau_d)|^2. \quad (31)$$

Expression (31) simplifies significantly in a few limiting cases:

(1) $\tau_{th} \ll \tau_p$:

$$\langle \tilde{\eta}(t) \rangle \propto \frac{\tau_c}{2\tau_{th}} + |\Gamma^{(2)}(\tau_d)|^2. \quad (32)$$

Thus, within a factor of 2, the thermal-grating decay time replaces the pulse length in expression (21). This makes sense because, recalling the argument leading to expressions (1) and (2), deposited energy now remains for a time, τ_{th} , only. As a result, we must now consider the number of coherence times in a thermal-grating decay time rather than in a pulse length, because grating-fringe contributions created more than τ_{th} ago are no longer present.

(2) $\tau_p \ll \tau_{th}$:

$$\langle \tilde{\eta}(t) \rangle \propto \frac{\tau_c}{\tau_p} + |\Gamma^{(2)}(\tau_d)|^2. \quad (33)$$

This result is just expression (21), as expected.

(3) $\tau_p < \tau_{th}$:

$$\langle \tilde{\eta}(t) \rangle \propto \frac{\tau_c}{\tau_p} \left[1 + \frac{1}{12} \left(\frac{\tau_p}{\tau_{th}} \right)^2 \right] + |\Gamma^{(2)}(\tau_d)|^2. \quad (34)$$

From this approximate result, we can see that the thermal-grating decay time must be as short as, or shorter than, the pulse length before decay effects are observed. The overall strength of the diffraction process depends sensitively on the thermal-grating decay time, but the ratio of the two terms, as indicated in expression (34), does not sensitively depend on τ_p/τ_{th} unless τ_{th} is less than τ_p .

7. INTEGRATED-INTENSITY GRATINGS INDUCED BY PARTIALLY COHERENT LIGHT FROM TWO SEPARATE LASERS

In the limit of large delays (specifically, $\tau_c \ll \tau_d$), the problem approximates that of using two separate, independent sources of the same frequency. We now specifically consider the problem of integrated-intensity-grating formation with two independent sources with potentially different frequencies. This problem is of interest because many researchers employing (two-excitation-laser) variable-fre-

quency, induced-grating techniques to measure excited-state relaxation times have instead seen only thermal-grating effects.^{9,12} An understanding of thermal gratings in two-excitation-laser experiments is important for the development of techniques for their suppression.²⁷

The simple argument at the beginning of Section 2 shows why a nonzero diffraction efficiency is to be expected when the frequency difference between the two excitation lasers, $\Delta\omega$, is zero. We now generalize and formalize that result for all values of $\Delta\omega$. Recalling Eq. (17) and observing that the two radiation fields, $u_1(t)$ and $u_2(t)$, will now be independent, we can separate the expectation operator into the product of two expectations. The expected normalized diffraction efficiency will now be

$$\langle \tilde{\eta}(\Delta\omega) \rangle = \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} \Gamma_1^{(2)}(t_1; t_2) \Gamma_2^{(2)*}(t_1; t_2) dt_1 dt_2, \quad (35)$$

where $\Gamma_i^{(2)}(t_j; t_k)$ is the second-order statistical coherence function of the i th excitation-beam radiation field.

Because the excitation lasers will, in general, lase at different frequencies, $\Gamma_1^{(2)}(t_1; t_2) \neq \Gamma_2^{(2)}(t_1; t_2)$, even if the lasers are otherwise identical. For wide-sense-stationary light sources, the arguments of the coherence functions can be written: $t_1 - t_2$. As a result,

$$\langle \tilde{\eta}(\Delta\omega) \rangle = \frac{1}{\tau_p^2} \int_{-\tau_p/2}^{\tau_p/2} \int_{-\tau_p/2}^{\tau_p/2} \Gamma_1^{(2)}(t_1 - t_2) \times |\Gamma_2^{(2)*}(t_1 - t_2)| dt_1 dt_2. \quad (36)$$

A routine change of variables ($t = t_1 - t_2$; $s = t_1 + t_2$) allows one integration to be performed trivially:

$$\langle \tilde{\eta}(\Delta\omega) \rangle = \frac{1}{2\tau_p} \int_{-\tau_p}^{\tau_p} (1 - |t/\tau_p|) \Gamma_1^{(2)}(t) \Gamma_2^{(2)*}(t) dt, \quad (37)$$

where $\tau_p' = \tau_p/\sqrt{2}$. Equation (37) can be evaluated exactly for the various statistical models we have been considering.

A general result can be obtained, however, in a useful limiting case. When each coherence time is much less than the pulse length, the functions $|\Gamma_i^{(2)}(t)|^2$ will be close to zero for all values of t for which the factor $(1 - |t/\tau_p|)$ deviates significantly from unity. This factor can thus be set equal to one with good accuracy. In addition, the integration limits may be extended to $-\infty$ and $+\infty$, respectively, in this limit. We then have, if $\tau_c \ll \tau_p$:

$$\langle \tilde{\eta}(\Delta\omega) \rangle \approx \frac{1}{\tau_p} \int_{-\infty}^{\infty} \Gamma_1^{(2)}(t) \Gamma_2^{(2)*}(t) dt, \quad (38)$$

and thus the thermal-grating line shape is just the Fourier transform of the product of the magnitudes of the second-order coherence functions of the excitation fields. Expression (38) can be rewritten in a different form using the Wiener-Khinchine theorem,¹¹ assuming wide-sense-stationary excitation fields:

$$\langle \tilde{\eta}(\Delta\omega) \rangle \approx \frac{1}{\tau_p} \int_{-\infty}^{\infty} I_1(\omega) I_2(\omega) d\omega, \quad (39)$$

which is just the convolution of the excitation frequency spectra: If we substitute into expression (38) or (39) the appropriate functions corresponding to Lorentzian and Gaussian line shapes, we obtain thermal-grating line shapes

for a Lorentzian line, in which $\tau_c = 2/\delta\omega$, and

$$\Gamma^{(2)}(t) = \exp(i\omega_0 t) \exp(-\pi t^2/2\tau_c^2) \quad (\text{A4})$$

for a Gaussian line, in which $\tau_c = \sqrt{8\pi \ln 2}/\delta\omega$. In each of the above expressions, τ_c is the radiation-field coherence time, which is necessarily less than the pulse length, τ_p , and which is defined so that

$$\tau_c = \int_{-\infty}^{\infty} |\Gamma^{(2)}(t)|^2 dt. \quad (\text{A5})$$

The diffraction-efficiency relations can be evaluated without the need for further approximations. We obtain, for an amplitude-stabilized quasi-monochromatic Lorentzian line,

$$\langle \tilde{\eta} \rangle = \left[\frac{\tau_c}{\tau_p} - \frac{1}{2} \frac{\tau_c^2}{\tau_p^2} \right] + \left[1 - 2 \frac{\tau_d}{\tau_p} + \frac{\tau_d^2}{\tau_p^2} + \frac{\tau_d \tau_c}{\tau_p^2} + \frac{1}{2} \frac{\tau_c^2}{\tau_p^2} - \frac{\tau_c}{\tau_p} \right] \exp(-2|\tau_d/\tau_c|), \quad (\text{A6})$$

which, to first order in τ_c/τ_p and τ_d/τ_p , is

$$\langle \tilde{\eta} \rangle \approx \frac{\tau_c}{\tau_p} + \left[1 - 2 \frac{\tau_d}{\tau_p} - \frac{\tau_c}{\tau_p} \right] \exp(-2|\tau_d/\tau_c|). \quad (\text{A7})$$

Other statistical models give similar results. For a thermal Lorentzian line, we obtain

$$\langle \tilde{\eta} \rangle = \left[\frac{\tau_c}{\tau_p} - \frac{1}{2} \frac{\tau_c^2}{\tau_p^2} \right] + \left[1 + \frac{\tau_c^2}{\tau_p^2} \right] \exp(-2|\tau_d/\tau_c|), \quad (\text{A8})$$

which, to first order in τ_c/τ_p , is

$$\langle \tilde{\eta} \rangle \approx \frac{\tau_c}{\tau_p} + \exp(-2|\tau_d/\tau_c|) \quad (\text{A9})$$

and which, in general, looks much like Eq. (A6). And for a thermal Gaussian line, we find

$$\langle \tilde{\eta} \rangle = \sqrt{2} \frac{\tau_c}{\tau_p} \operatorname{erf} \left(\sqrt{2} \frac{\tau_p}{\tau_c} \right) + \exp(-\pi \tau_d^2/\tau_c^2) - \frac{1}{\pi} \frac{\tau_c^2}{\tau_p^2} [1 - \exp(-\pi \tau_p^2/\tau_c^2)], \quad (\text{A10})$$

which, to first order in τ_c/τ_p , becomes

$$\langle \tilde{\eta} \rangle \approx \sqrt{2} \frac{\tau_c}{\tau_p} + \exp(-\pi \tau_d^2/\tau_c^2), \quad (\text{A11})$$

which also looks similar to the other curves. The Gaussian shape deviates, of course, from the simple exponentials obtained for the Lorentzian lines, but the limiting cases are as expected, with the $\sqrt{2}$ having its origins in the definitions of τ_c and τ_p .

APPENDIX B: THE INCLUSION OF PULSE-SHAPE EFFECTS

We will assume a Gaussian pulse shape:

$$A(t) = \left(\frac{4 \ln 2}{\pi \tau_p} \right)^{1/4} \exp[-2 \ln 2 (t^2/\tau_p^2)], \quad (\text{B1})$$

where τ_p is the FWHM of the intensity. $A(t)$ is normalized so that

$$\int_{-\infty}^{\infty} A^4(t) dt = \int_{-\infty}^{\infty} I^2(t) dt = 1, \quad (\text{B2})$$

i.e., a dimensionless, unity-magnitude autocorrelation.

For a thermal Lorentzian line and a Gaussian pulse shape then, Eq. (22) becomes

$$\langle \tilde{\eta} \rangle = \exp[-2 \ln 2 (\tau_d^2/\tau_p^2)] \left\{ \exp \left(\frac{\tau_p^2/\tau_c^2}{2 \ln 2} \right) \times \left[1 - \operatorname{erf} \left(\frac{\tau_p/\tau_c}{\sqrt{2 \ln 2}} \right) \right] + \exp(-2|\tau_d/\tau_c|) \right\}. \quad (\text{B3})$$

When $\tau_c \ll \tau_p$, this result simplifies considerably:

$$\langle \tilde{\eta} \rangle \approx \sqrt{\frac{2 \ln 2}{\pi}} \left(\frac{\tau_c}{\tau_p} \right) \exp[-2 \ln 2 (\tau_d^2/\tau_p^2)] + \exp(-2|\tau_d/\tau_c|). \quad (\text{B4})$$

A thermal Gaussian line with a Gaussian pulse shape yields

$$\langle \tilde{\eta} \rangle = \exp[-2 \ln 2 (\tau_d^2/\tau_p^2)] \left\{ \sqrt{\frac{2 \ln 2}{\pi}} \left(\frac{\tau_c}{\tau_p} \right) \times \left(1 + \frac{2 \ln 2}{\pi} \frac{\tau_c^2}{\tau_p^2} \right)^{-1/2} + \exp(-\pi \tau_d^2/\tau_c^2) \right\}, \quad (\text{B5})$$

which also simplifies when $\tau_c \ll \tau_p$:

$$\langle \tilde{\eta} \rangle \approx \sqrt{\frac{2 \ln 2}{\pi}} \left(\frac{\tau_c}{\tau_p} \right) \exp[-2 \ln 2 (\tau_d^2/\tau_p^2)] + \exp(-\pi \tau_d^2/\tau_c^2). \quad (\text{B6})$$

The fourth-order theory for a thermal Lorentzian line using a hyperbolic-secant-squared pulse shape results in a diffraction efficiency:

$$\langle \tilde{\eta} \rangle \approx \frac{\zeta \tau_c}{\tau_p} \left[\frac{\zeta \tau_d/\tau_p \cosh(\zeta \tau_d/\tau_p) - \sinh(\zeta \tau_d/\tau_p)}{\sinh^3(\zeta \tau_d/\tau_p)} \right] + \exp(-2|\tau_d/\tau_c|), \quad (\text{B7})$$

where $\zeta = 1.7627$.

ACKNOWLEDGMENTS

R. Trebino would like to acknowledge the helpful conversations with and advice of J. W. Goodman. The authors are grateful for financial assistance from the U.S. Air Force Office of Scientific Research.

REFERENCES AND NOTES

1. G. Martin and R. W. Hellwarth, "Infrared-to-optical image conversion by Bragg reflection from thermally induced index changes," *Appl. Phys. Lett.* **34**, 371-373 (1979).
2. R. K. Jain and M. B. Klein, "Degenerate four-wave mixing in semiconductors," in *Optical Phase Conjugation*, R. A. Fisher, ed. (Academic, New York, 1983), pp. 365-366.
3. J. R. Salcedo, A. E. Siegman, D. D. Dlott, and M. D. Fayer, "Dynamics of energy transport in molecular crystals: the picosecond transient grating method," *Phys. Rev. Lett.* **41**, 131-134 (1978).
4. K. A. Nelson, D. R. Lutz, M. D. Fayer, and L. Madison, "Laser-induced phonon spectroscopy, optical generation of ultrasonic

- waves and investigation of electronic excited-state interactions in solids," *Phys. Rev. B* **24**, 3261-3275, (1981).
5. J. R. Salcedo and A. E. Siegman, "Laser induced photoacoustic grating effects in molecular crystals," *IEEE J. Quant. Electron.* **QE-15**, 250-256 (1979).
 6. V. S. Idiutulin and Yu. N. Teryaev, "Transient gratings in a nonlinear medium," *Opt. Quantum Electron.* **14**, 51-56 (1982).
 7. H. J. Eichler, U. Klein, and D. Langhans, "Coherence time measurement of picosecond pulses by a light-induced grating method," *Appl. Phys.* **21**, 215-219 (1980).
 8. R. Baltrameynas, Yu. Vaitkus, R. Dannelyus, M. Pyatrauskas, and A. Piskarskas, "Applications of dynamic holography in determination of coherence times of single picosecond light pulses," *Sov. J. Quantum Electron.* **12**, 1252-1254 (1982).
 9. J. R. Andrews and R. M. Hochstrasser, "Transient grating studies of energy deposition in radiationless processes," *Chem. Phys. Lett.* **76**, 207-212, (1980).
 10. Z. Vardeny and J. Tauc, "Picosecond coherence coupling in the pump and probe technique," *Opt. Commun.* **39**, 396-400 (1981).
 11. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
 12. R. Trebino, "Subpicosecond-relaxation studies using tunable-laser-induced-grating techniques," Ph.D. dissertation (Stanford University, Stanford, Calif., 1983).
 13. B. S. Wherrett, A. L. Smirl, and T. F. Boggess, "Theory of degenerate four-wave mixing in picosecond excitation-probe experiments," *IEEE J. Quantum Electron.* **QE-19**, 680-689 (1983).
 14. C. V. Shank and D. H. Auston, "Parametric coupling in an optically excited plasma in Ge," *Phys. Rev. Lett.* **34**, 479 (1975).
 15. A. von Jena and H. E. Lessing, "Coherent coupling effects in picosecond absorption experiments," *Appl. Phys.* **19**, 131-144 (1979).
 16. T. F. Heinz, S. L. Palfrey, and K. B. Eisenthal, "Coherent coupling effects in pump-probe measurements with collinear, copropagating beams," *Opt. Lett.* **9**, 359-361 (1984).
 17. W. M. Grossman and D. M. Shemwell, "Coherence lengths and phase conjugation by degenerate four-wave mixing," *J. Appl. Phys.* **51**, 914-916. (1980).
 18. M. Schubert, K.-E. Süssé, and W. Vogel, "Influence of chaotic pump radiation with finite bandwidth on the intensity correlation of resonance fluorescence radiation," *Opt. Commun.* **30**, 275-278 (1979).
 19. H. J. Kimble and L. Mandel, "Resonance fluorescence with excitation of finite bandwidth," *Phys. Rev. A* **15**, 689-699 (1977).
 20. It appears, from our discussion, that in the cw limit the diffraction efficiencies in both cases go to infinity. This is, of course, not the case because material relaxation prevents the buildup of infinite grating strength. When the pulse is longer than the material relaxation time, the relaxation time then replaces the pulse length in the above argument. Section 6 treats this effect in greater detail. In any event, the argument presented in the text is intended to give relative grating strengths only.
 21. Equations (11) and (12) explain why thermal gratings have obscured population gratings in grating experiments on dyes dissolved in ethanol but not in experiments using water as a solvent. The solvent-dependent factors in Eq. (11) result in a thermal-grating diffraction efficiency proportional to $[(dn/dT)/\rho c_v]^2$, with other factors depending on the solute, laser light, or beam geometry or exhibiting little variation.⁹ The value of $[(dn/dT)/\rho c_v]^2$ for ethanol is more than 100 times that for water.
 22. R. C. Desai, M. D. Levenson, and J. A. Baker, "Forced Rayleigh scattering: thermal and acoustic effects in phase-conjugate wave-front generation," *Phys. Rev. A* **27**, 1968-1976 (1983).
 23. A. E. Siegman, "Bragg diffraction of a Gaussian beam by a crossed-Gaussian volume grating," *J. Opt. Soc. Am.* **67**, 545-550 (1955).
 24. B. Picinbono and E. Boileau, "Higher-order coherence functions of optical fields and phase fluctuations," *J. Opt. Soc. Am.* **58**, 784-789 (1968).
 25. J. J. Song, J. H. Lee, and M. D. Levenson, "Picosecond relaxation measurements by polarization spectroscopy in condensed phases," *Phys. Rev. A* **17**, 1439-1447 (1978).
 26. J. R. Salcedo, A. E. Siegman, D. D. Dlott, and M. D. Fayer, "Dynamics of energy transport in molecular crystals: the picosecond transient grating method," *Phys. Rev. Lett.* **41**, 131-134 (1978).
 27. R. Trebino, A. E. Siegman, and C. L. Ladera, "Suppression of thermal gratings in polarization spectroscopy," *J. Opt. Soc. Am. B* **1**, 549-550 (1984).

FREQUENCY BANDWIDTHS IN NONDEGENERATE N-WAVE-MIXING INTERACTIONS AND INDUCED-GRATING GEOMETRIES

Rick TREBINO and A.E. SIEGMAN

Department of Applied Physics, Stanford University, Stanford, CA 94305, USA

Received 13 August 1985

We calculate frequency bandwidths for variable-frequency *N*-wave-mixing interactions and show that the most important parameter for bandwidth determinations is the angle between the two beams (most often one input and one output beam) whose frequencies are varying. Geometries in which the two variable-frequency beams copropagate attain the largest bandwidths, while counterpropagating variable-frequency beams produce the narrowest bandwidths. When more than two beams vary in frequency, additional angles are necessary to characterize the interaction: we calculate bandwidths for the planar case and discuss autocorrelators employing second harmonic generation. Finally, we briefly discuss the merits of a few broadband geometries.

1. Introduction

Variable-frequency *N*-wave-mixing interactions and induced-grating methods form the basis of a rich variety of techniques and devices. Saturation spectroscopy [1,2], polarization spectroscopy [3,4], coherent anti-Stokes Raman spectroscopy [5], and the tunable-laser-induced-grating technique [6–8] are examples of such processes for the case $N = 4$. Pulselength-measurement techniques based on second-harmonic-generation autocorrelators [9], when applied to picosecond tunable dye lasers, are variable-frequency three-wave-mixing processes.

In utilizing such interactions, it is usually important to attain a large frequency bandwidth, that is, to maintain phase-matching despite large variations in two or more of the input or output beam frequencies [8]. Researchers have employed a wide variety of geometries for *N*-wave-mixing interactions and induced-grating experiments, and various authors have treated the bandwidth problem in the past, usually reporting the discovery of a particularly good (broadband) or bad (narrowband) geometry [10–15]. (Of course, if one is interested in optical frequency filters [11], the good-bad dichotomy reverses.) It is well-known, for example, that the four-wave-mixing phase

conjugate geometry is a particularly narrowband geometry [10,16].

Although such specific results can be found in the literature, a general treatment of the geometry-dependent bandwidth issue does not appear to exist. Consequently, it is the purpose of this paper to present such a treatment – which will be seen to be quite simple. We also give an intuitive picture for these results. We begin with two sections treating the case in which only two frequencies vary and then include a section on the more complicated situation involving three variable frequencies. A discussion of broadband wave-mixing geometries completes the analysis.

2. Interactions producing a variable-frequency output beam

We first consider the two-variable-frequency case in which one input frequency and the output frequency vary. Our treatment will suffice for all *N*-wave-mixing (and induced-grating) interactions, although it was first presented to explain the narrowband nature of the four-wave-mixing phase-conjugate process, in which the input and the counterpropagating output beam vary in frequency.

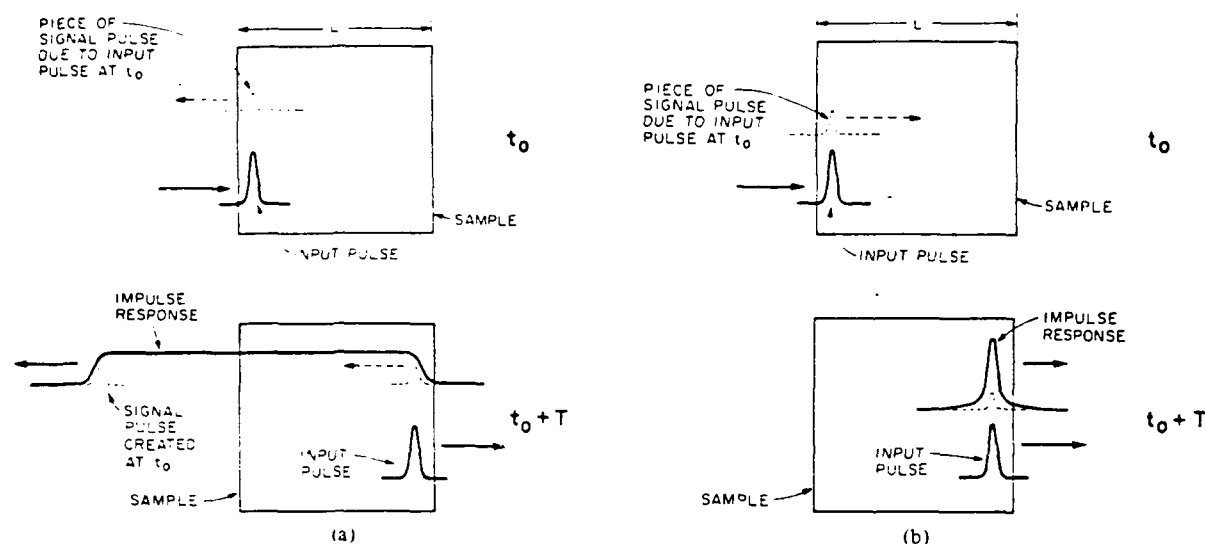


Fig. 1. (a) Illustration of the finite time response, and hence narrow bandwidth, of an interaction involving counterpropagating output and variable-frequency input beams. The box corresponds to the interaction region. The upper pulse in each box is in part of the output pulse, created by the interaction, while the lower pulse in each box is the input pulse; their vertical separation is for the purposes of illustration only. (b) Illustration of the instantaneous time response and hence, very large bandwidth, of an interaction involving copropagating output and variable-frequency input beams. See comments in the caption of (a).

Following Siegman et al. [16], suppose that the wave-mixing process has an interaction length L , and a corresponding transit time $T = L/c$ along the output-wave direction (fig. 1a). Let the fixed-frequency beams be cw waves and the variable-frequency input beam be a short pulse (with pulse-length much less than T). The wave-mixing process will create output radiation only at points in time and space when all input beams coincide. As a result, when the input pulse *counterpropagates* with respect to the about-to-be-created output wave, the short pulse will shed a continuous "tail" of induced output wave. The output wave begins to appear at the instant that the input pulse first arrives at the entrance to the interaction region, but radiation created later cannot "catch up" with radiation created earlier: the younger radiation experiences later birth, but in addition, its point of creation is *behind* the older output radiation. As a result, the approximate pulse-shape of the output wave will be a square wave of duration $2T$ or $2L/c$. The time response of the interaction is thus not instantaneous.

The frequency response of the interaction, if one

considers this as a linear system with input field, is then the Fourier transform of the above impulse response. Because the impulse response is broadened by an amount on the order of $2T$ in time, the frequency response of the interaction with respect to the variable-frequency input beam will be finite, that is, the geometry will have a narrow bandwidth. In other words, frequency scans broad compared to $1/2T$, a large phase-mismatch will occur (or equivalently, the Bragg angle for the interaction will vary from its original value or cease to exist).

Now consider a geometry in which the output beam and variable-frequency input beam *copropagate* (fig. 1b). We can calculate the impulse response for this interaction in an analogous manner: it is easy to see that the impulse response for such a geometry is a delta-function. Because the output beam and input beam copropagate, the output beam will consist of an identical short pulse copropagating — both in space and in time — with the short input pulse. The frequency response of this system will be large, and hence, despite a large frequency scan, very little or no phase mismatch will result.

We have neglected dispersion in this argument. The presence of dispersion will of course weaken this argument somewhat: if the two pulses have different group velocities, they will separate as they propagate, and the impulse response will actually have nonzero width. However, its width will in general still be much less than that of the counterpropagating-beam case.

Thus, wave-mixing geometries in which the output beam and the variable-frequency input beam *counter*-propagate will be narrowband, and, conversely, geometries in which these beams *co*propagate will be broadband. In particular, four-wave-mixing phase conjugators (in which these beams counterpropagate) have narrow bandwidths, as is well-known. What is not so well-known, however, is that such phase conjugators have narrow bandwidth *because* these beams counterpropagate.

3. Interactions involving two arbitrary variable-frequency beams

The argument in the preceding section demonstrates that wave-mixing geometries in which one input frequency and the output frequency vary will have a maximally broad bandwidth if the variable-frequency beams copropagate. We now extend this result to *N*-wave-mixing geometries in which *any* two frequencies vary. We also work more rigorously, deriving an analytical expression for the interaction bandwidth.

Consider a general *N*-wave mixing process for which

$$\sum_{i=1}^N \mu_i \omega_i = 0, \quad (1)$$

where $\mu_i = \pm 1$, and ω_i is a (positive-valued) input or output frequency. We make no distinction between input and output beams since they enter equivalently into the analysis. If the *k*-vectors of the interaction are labeled k_i for $i = 1 \dots N$, the phase-mismatch will be

$$\Delta k = \sum_{i=1}^N \mu_i k_i. \quad (2)$$

Now suppose that initially the process is phasematched ($\Delta k = 0$), but that two frequencies are varied while

all beam directions remain fixed. Necessarily, eq. (1) will remain satisfied, so that if ω_i is incremented by $\delta\omega_i$ and ω_j by $\delta\omega_j$, we have

$$\mu_i \delta\omega_i + \mu_j \delta\omega_j = 0. \quad (3)$$

If Ω is the angle between the two variable-frequency beams, the squared magnitude of the phase-mismatch due to the variation in frequencies will be

$$|\Delta k|^2 = (\delta\omega_i^2/c^2)(n_i^2 + n_j^2 - 2n_i n_j \cos \Omega) \quad (4)$$

neglecting dispersion. Setting $|\Delta k|^2$ equal to $(\pi/L)^2$, where L is the interaction length, allows simple solution for the actual bandwidth, $\Delta\omega_i$

$$\Delta\omega_i = (2\pi c/L)[(n_i - n_j)^2 + 4n_i n_j \sin^2(\Omega/2)]^{-1/2}, \quad (5)$$

which is plotted in fig. 2. We see that the bandwidth will be maximized when $\Omega = 0$. It is, when the variable-frequency beams *co*propagate. The optimal case, $\Delta\omega_i \rightarrow \infty$, occurs when these beams copropagate and their refractive indices are equal: $n_i = n_j$. When the refractive indices are unequal, the bandwidth will decrease for all geometries, but copropagating beams will continue to maximize it. In addition, note that the bandwidth is independent of μ_i and μ_j . Finally, for degenerate processes involving m_i photons from the beam of frequency ω_i and *k*-vector k_i , the bandwidth will be reduced by a factor of m_i ; all other conclusions remain valid.

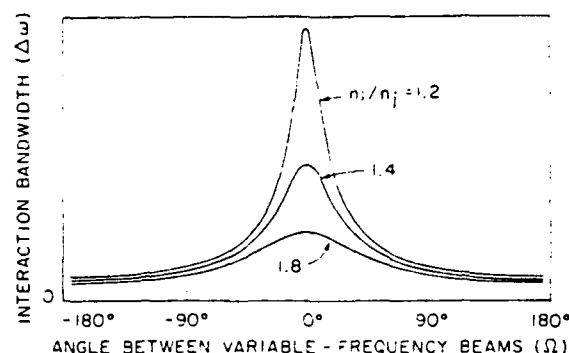


Fig. 2. The interaction bandwidth for an *N*-wave-mixing interaction in which two beam frequencies vary versus the angle between these two beams (Ω). The minimum bandwidth, occurring when $\Omega = \pm\pi$, is given by: $\Delta\omega_{\min} = 2\pi c/[(n_i + n_j)L]$, while the maximum bandwidth attainable by a process, occurring for copropagating variable-frequency beams ($\Omega = 0$) is given by $\Delta\omega_{\max} = 2\pi c/[(n_i - n_j)L]$, or if $n_i = n_j$, $\Delta\omega_{\max} = 2\pi c/(\omega_i n_i - \omega_j n_j / L)$. When three frequencies vary, copropagation may not yield the maximal bandwidth.

Dispersive effects generally play only a very small role in the determination of the bandwidth since index-mismatch effects due to the changes in frequency usually predominate. Actual calculation of the geometry bandwidth will, however, occasionally require the additional dispersive terms, since choosing a large-bandwidth geometry often forces cancellation or near-cancellation of the refractive-index terms in eq. (5), with the dispersive terms then dominating. We can include dispersion effects in this analysis by replacing $n_\eta + \omega_\eta n'_\eta$, where $n'_\eta = \partial n / \partial \omega|_{\omega=\omega_\eta}$ and $\eta = i, j$. Copropagating beams will continue to maximize the bandwidth provided that $|\omega_\eta n'_\eta| < n_\eta$.

4. Interactions with three variable-frequency beams

We can generalize eq. (4) to N -wave-mixing processes in which *three* frequencies (ω_i , ω_j and ω_l) vary, obtaining

$$\begin{aligned} |\Delta k|^2 = & (\delta\omega_i^2/c^2)[n_i^2 + n_l^2 - 2n_i n_l \cos \Omega_{il}] \\ & + (\delta\omega_j^2/c^2)[n_j^2 + n_l^2 - 2n_j n_l \cos(\Omega_{ij} - \Omega_{il})] \\ & + 2\mu_i \mu_j (\delta\omega_i \delta\omega_j / c^2)[n_i^2 + n_j n_l \cos \Omega_{ij} \\ & - n_i n_l \cos \Omega_{il} - n_j n_l \cos(\Omega_{ij} - \Omega_{il})], \end{aligned} \quad (6)$$

where Ω_{ij} is the angle between k_i and k_j , and now, Ω_{il} is the angle between k_i and k_l measured in the same direction as Ω_{ij} . We assume a coplanar arrangement. Notice that now μ_i and μ_j appear and may help to determine the optimal values of Ω_{ij} and Ω_{il} for a particular process. The bandwidth for a particular interaction will depend on the relation between $\delta\omega_i$ and $\delta\omega_j$, and is easily calculated from eq. (6) upon appropriate substitution.

Application of this result to specific wave-mixing processes is straightforward, although computationally messy. For example, consider a second-harmonic-generation (SHG) autocorrelator (e.g., for ultrashort-pulsewidth measurement of dye-laser pulses), for which it is desirable to minimize $|\Delta k|$, and hence the amount of realignment necessary, as the input frequency is varied. Since these devices generally employ a non-collinear-input-beam geometry, we require the use of eq. (6). Using $\delta\omega_i = \delta\omega_j = \delta\omega_l/2$ and $\mu_i = \mu_j = -\mu_l$, we find

$$\cos \Omega_{il} = (n_i^2 - n_j^2 + 4n_l^2)/4n_i n_l, \quad (7)$$

$$\sin \Omega_{ij} = (2n_l/n_j) \sin \Omega_{il}, \quad (8)$$

for the beam angles that optimize the autocorrelator bandwidth. Since $n_l = (n_i + n_j)/2$ for an initially phase-matched interaction, we find that $\Omega_{ij} = \Omega_{il} = 0$, and again, copropagating beams optimize the bandwidth. In practice, most commercial autocorrelators [9] employ non-zero values of Ω_{ij} and Ω_{il} in order to obtain background-free operation, thus sacrificing some bandwidth.

In a higher-order process with one or more fixed-frequency beams in addition to the variable-frequency beams considered above for the autocorrelator, we obtain a bandwidth-optimizing beam geometry with *noncollinear* variable-frequency beams. Additional beam(s) allow a more general index-matching equation, which we will now take to be: $n_l = (n_i + n_j - \delta)/2$, where δ incorporates all fixed-frequency-beam refractive indices. It is then easy to show that, for small values of δ , the optimal beam angles, Ω_{ij} and Ω_{il} will be

$$\begin{aligned} \Omega_{il} & \approx [(n_j/n_i n_l) \delta]^{1/2} & \text{if } \delta \geq 0, \\ & \approx 0 & \text{if } \delta < 0, \end{aligned} \quad (9)$$

$$\begin{aligned} \Omega_{ij} & \approx 2[(n_l/n_i n_j) \delta]^{1/2} & \text{if } \delta \geq 0, \\ & \approx 0 & \text{if } \delta < 0, \end{aligned} \quad (10)$$

working to first order in δ . Thus, collinear propagation of the variable-frequency beams is not the universal solution to the problem of N -wave-mixing bandwidth maximization when more than two beams vary in frequency.

5. Discussion and conclusions

While it is well known that phase conjugation by four-wave mixing is an inherently narrowband process^{*}, the set of broadband four-wave-mixing geometries is not so well-known. In this section, we discuss a few geometries that achieve broad bandwidth and, hence, are appropriate for variable-frequency non-

^{*}For footnote see next page.

Fig. polarized frequency beams propagating in the same direction as the beams.

linear polarization.

specific frequency.

* For footnote see next page.

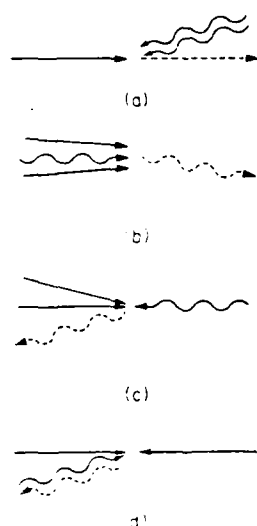


Fig. 3. (a) The infinite-bandwidth geometry of saturation and polarization spectroscopy. The wiggly lines denote variable-frequency beams, and the dashed line represents the output beam. (b) The large-bandwidth geometry involving near-copropagation of all beams. The wiggly lines denote variable-frequency beams, and the dashed line represents the output beam. (c) Another large-bandwidth geometry. The wiggly lines denote variable-frequency beams, and the dashed line represents the output beam. In this case, two beams can be made to counterpropagate [7,8], thus simplifying alignment considerably. See text for additional advantages and disadvantages of this geometry. (d) The (narrow-bandwidth) geometry of phase conjugation by four-wave mixing. The wiggly lines denote variable-frequency beams, and the dashed line represents the output beam. Observe that the variable-frequency beams *counterpropagate*.

linear-optical techniques. We also consider the practical problems of the presence of backgrounds, such as polarizer leakage and small-angle scattered light.

The usual geometry for saturation and polarization spectroscopy [1-4] (fig. 3a) employs two variable-frequency beams combined in the same beam emanating from the same laser. The output beam remains at

a fixed frequency. Like phase conjugation, these interactions, when viewed as four-wave-mixing processes, are automatically phase-matched. Unlike phase conjugation, however, these processes achieve *infinite* bandwidth due to the necessary exact copropagation and the same wavelengths and indices of refraction of the variable-frequency beams. This conclusion is, of course, independent of dispersion effects. On the negative side, however, this geometry always possesses a background of at least polarizer leakage.

The use of a fixed-frequency output beam in the above geometry requires the variation of two input frequencies by exactly the same amounts. This is easy if the beams are one and the same, but prohibitively difficult otherwise. Consequently, most variable-frequency wave-mixing experiments control a single input-beam frequency allowing the output-beam frequency to vary as a result. A popular geometry of this sort is the collinear-beam geometry, which attains very long interaction lengths, but possesses a beam separation problem when the wavelengths involved are approximately equal. *Near-copropagation* of all beams (fig. 3b) is a useful alternative arrangement that allows easier beam separation [18]. Still, a problem exists in the form of small-angle-scattered-light background from all of the input beams, which can be reduced by using larger angles between the beams, but at a cost in frequency bandwidth.

As a partial solution to this trade-off, we have employed a broadband four-wave-mixing geometry (fig. 3c) in which the two variable-frequency beams approximately copropagate, but the other two beams emanate from other angles. Such a geometry can significantly reduce the scattered light in a given beam direction, reducing the number of beams producing scattered light to a maximum of one. In earlier experiments [7] we allowed the other two (fixed-frequency) beams to nearly counterpropagate with the variable-frequency beams to maintain a long interaction length and ease of alignment. (Polarization- and wavelength-filtering helped to reduce the scattered-light background from the one beam that remained in a nearly copropagating direction with the output beam.) This type of four-wave-mixing geometry is useful, and we obtained good results with it, but it should be mentioned here that, unlike the other geometries mentioned in this section, this arrangement suffers from its own drawbacks. It has severe constraints regarding the

* "Forward-going phase conjugation" [17] employs the variable-frequency beams in a nearly copropagating geometry, and hence, achieves the desired large bandwidth. Such an arrangement has its drawbacks, however: mainly a small *angular* bandwidth [17] in contrast to the 2π angular bandwidth of the usual "backward-going" phase conjugation. As a result, researchers generally shun forward-going phase conjugation, since angular bandwidth is more important in the phase-conjugation process than frequency bandwidth.

allowable wavelengths in the interaction: often, the wavelengths involved cause phasematching and a long interaction length to be incompatible due to the required Bragg angle, and, for some wavelengths, phasematching is impossible. Ref. [8] provides a more complete analysis of this problem and of four-wave mixing geometries, in general.

We must conclude that there is no wave-mixing geometry that is ideal for all applications requiring broad bandwidth. Choice of appropriate geometry will always require thoughtful analysis of the material parameters and sources of experimental noise, in particular. In any case, an understanding of the significant role played by the relevant beam angles of the geometry in determining the frequency bandwidth of the interaction is certainly helpful.

Acknowledgement

The authors would like to thank David Ziony for graphical assistance and Adnah Kostenbauder for a critical reading of this manuscript. The authors would also like to acknowledge the financial support of the Air Force Office of Scientific Research.

References

- [1] M. Sargent III, *Appl. Phys.* 9 (1976) 127.
- [2] F. Keilmann, *IEEE J. Quant. Electron.* QE-12 (1976) 592.
- [3] J.J. Song, J.H. Lee and M.D. Levenson, *Phys. Rev. A* 17 (1978) 1439.
- [4] J.J. Lee, J.J. Song, M.A.F. Scarparo and M.D. Levenson, *Optics Lett.* 5 (1980) 196.
- [5] A. Compaan and S. Chandra, *Optics Lett.* 4 (1979) 170.
- [6] A.E. Siegman, *Appl. Phys. Lett.* 30 (1977) 21.
- [7] R. Trebino and A.E. Siegman, *J. Chem. Phys.* 79 (1983) 3621.
- [8] P. Trebino, *Subpicosecond-relaxation studies using tunable-laser-induced-grating techniques*. Ph.D. dissertation, Stanford University, May 1983.
- [9] Spectra Physics; Coherent, Inc.
- [10] R.L. Abrams, J.F. Lam, R.C. Lind and D.G. Steel, in: *Optical phase conjugation* (Academic Press, New York, 1984) pp. 211-284.
- [11] D.M. Pepper and R.L. Abrams, *Optics Lett.* 3 (1978) 212.
- [12] A.C. Eckbrecht, *Appl. Phys. Lett.* 32 (1978) 421.
- [13] G. Laufer and R.B. Miles, *Optics Comm.* 28 (1979) 250.
- [14] E.J. Heilweil, R.M. Hochstrasser and H. Souma, *Optics Comm.* 35 (1980) 227.
- [15] S. Chandra, A. Compaan and E. Wiener-Avnear, *Appl. Phys. Lett.* 33 (1978) 867.
- [16] A.E. Siegman, P.A. Belanger and A. Hardy, in: *Optical phase conjugation* (Academic Press, New York, 1983) pp. 465-535.
- [17] C.V. Heer and N.C. Griffen, *Optics Lett.* 4 (1979) 239.
- [18] T. Yajima and H. Souma, *Phys. Rev. A* 17 (1978) 309.

Cor
cepted
ment
in Eng
(plus
all pro
should
4. x
margin
must
contai
be let
width

For p

1. P
of

2. P
h.

3. O
ti.

4. O
5. A.

6. C
7. P

8. I
9. N

In t

Achromatic N-prism beam expanders: optimal configurations II

Rick Trebino, Charles E. Barker, and A. E. Siegman

Edward L. Ginzton Laboratory, Stanford University
 Stanford, California 94305

Abstract

We extend the analysis of Trebino (Appl. Opt. 24, Apr. 15, 1985) of single-material achromatic prism beam expanders to include two-material devices and practical expanders consisting entirely of commercially available 45-45-90 prisms. We show that sacrificing simultaneous achromaticity and thermal stability allows the construction of two-material prism beam expanders with only one of these two properties, but having significantly greater transmission than a similar-magnification single-material device. Finally, a numerical optimization routine yields optimal incidence angles for two-, three-, and four-prism single-material achromatic prism beam expanders. In general, we find that the "up-up-...-up-down" configuration is optimal in nearly all practical cases.

Introduction

When Klauminzer¹ introduced a prism beam expander into a pulsed dye laser in 1977, a century-old concept experienced a rebirth and has since found numerous applications in laser technology.² In particular, achromatic prism beam expanders have become important in broad-band applications and arrangements requiring high wavelength stability.³ The prism configuration, that is, which prism dispersions add (i.e., point "up") and which prism dispersions subtract (i.e., point "down"), while irrelevant in non-achromatic devices, is, of course, an important consideration in achromatic prism beam expanders. In this paper, we will review the results of Trebino,⁴ that is, that the optimal configuration of prisms in an achromatic prism beam expander depends explicitly on the desired magnification, and that, for single-material devices, the up-up-...-up-down configuration⁵ is optimal for most practical magnifications. The approach is to define a "characteristic polynomial" for each prism configuration and to show that the zeros of these polynomials correspond to exactly optimal (i.e., maximal transmission) achromatic prism beam expanders. For the single-material devices considered by Trebino, these solutions correspond to relatively low magnifications; consequently, here, we consider multi-material N-prism beam expanders. We show that, if the final prism of a single-material device is replaced by a prism of lower-dispersion material, exactly optimal achromatic devices of much higher magnification emerge from the analysis. In other words, such two-material devices achieve higher transmission. Finally, since these exactly optimal expanders generally employ odd, and, hence, inconvenient apex angles, we report here numerical results for optimal achromatic two-, three-, and four-prism beam expanders using only 45-45-90° prisms of a single material.

Optimal solutions: a review

For a not-necessarily-achromatic N-prism beam expander, Trebino,⁴ has solved exactly the general problem of determining the incidence and apex angles of the various arbitrary-material prisms, under the assumptions of uncoated entrance faces and AR-coated exit faces. In particular, if all the prism materials are identical, all of the prisms should be identical, with identical incidence angles, also. It follows that each prism's magnification should be identical. By assuming equal magnifications then for each prism in the equation for the prism-beam-expander dispersion, a "characteristic polynomial," $p(x)$, for a given prism configuration can be easily obtained:⁴

$$p(m) = \sum_{i=1}^N \epsilon_i m^{i-1} \quad (1)$$

where ϵ_i is the orientation of the i th prism (+1 for "up"; -1 for "down") and m is the magnification per prism for which this configuration happens to have the same transmission as the optimal not-necessarily-achromatic device. The zeros of Equation (1) are thus the magnifications per prism for which the $\{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$ configuration coincides with the already-proven optimal not-necessarily-achromatic prism beam expander. Obviously, these solutions thus represent optimal achromatic prism beam expanders of magnification $M = m^N$. Table 1 lists these solutions for single-material devices, and Figure 1 indicates the estimated regions about these values for which a given configuration is also optimal.

Table 1. Exact Optimal Achromatic Solutions for Single-Material Prism Beam Expanders

Number of Prisms (N)	Configuration	m	$M = m^N$	T
3	uud	1.62	4.42	1.00
4	uuud	1.84	11.44	0.96
	duud	1.00	1.00	0.85
	udud			
	uudd			
5	uuuud	1.93	26.61	0.92
	duuud	1.72	15.15	0.98
	uduud	1.51	7.93	1.00
	uudud	1.29	3.58	0.97
	uuudd	1.18	2.28	0.93
	several			
6	uuuuud	1.97	57.73	0.89
	duuuud	1.88	44.60	0.93
	uduud	1.79	33.16	0.95
	uuuud	1.62	17.94	0.99
	duuud	1.41	7.78	0.99
	uduud	1.27	4.24	0.96
	uuudd	1.00	1.00	0.78
	duudd			
	duuud			
	several			

Two-material achromatic prism beam expanders

It should be evident from these figures that the derived optimal solutions—even the highest-magnification, up-up-...-up-down configuration results—represent relatively low magnifications compared to those employed in typical applications. The reason for this is that, even if only the final prism's dispersion subtracts while all the others add, the final prism's dispersion tends to dominate at high magnifications because each prism's dispersion contribution to the total device dispersion is weighted by the inverse of the total magnification after that prism.⁴ A solution to this problem is relatively obvious: replace the final prism with one of lower-dispersion material. Higher-magnification solutions will, in fact, result, with the advantage, of course, that higher transmissions for a given magnification will also result.

This gain is not without its price, however. Single-material prism beam expanders have the advantage that achromaticity and thermal stability are simultaneously obtained independent of all material parameters (except for the refractive index).⁴ The use of more than one material in a prism beam expander allows only one of the above qualities in an expander—when both are often simultaneously desired.

With the above caveat in mind, we can proceed to construct optimal two-material prism beam expanders using a final prism of, say, one half of the dispersion of that of the other prisms. (If, instead, a thermally stable device is desired, the appropriate quantity is the derivative of the refractive index with respect to temperature rather than wavelength.) A simple manipulation shows that we must now replace ϵ_N in the characteristic polynomial by $\epsilon_N (dn_N/d\lambda)/(dn/d\lambda)$, where n_N is the refractive index of the final prism. (We are assuming equal refractive indices at the relevant wavelength here for simplicity.) Thus, the coefficients of the characteristic polynomial are not necessarily +1 or -1 in multi-material devices, and in general, these coefficients may take on any real value for more complex devices.

Table 2 lists the zeros to the characteristic polynomials for configurations with values greater than unity for two-material devices in which the final prism's dispersion is one half that of the other prisms. Similarly, Figure 2 indicates the estimated regions about these values for which a given configuration is probably also optimal. Observe that the optimal magnifications are significantly greater than those in the single-material case. Clearly, optimality of these solutions proves that the transmissions of these two-material devices will exceed those of single-material achromatic devices of the same magnification, although our formalism gives no indication as to the actual difference in transmissions of these devices. In general, however, this formalism yields a good beginning in the design of these possibly quite complex devices, using only the computing power of a hand-held calculator (the SOLVE routine of an HP-15C calculator provided the numbers in Tables 1 and 2).

Table 2. Exact optimal achromatic solutions for two-material prism beam expanders constructed in the following manner: the first $N-1$ prisms are of the same material, with the final prism having a value of $\partial n/\partial x$ equal to half that of the other prisms. For simplicity, all prisms are assumed to have the same refractive index. (All transmissions shown in Tables 1 and 2 assume $n = 1.5$).

Number of Prisms (N)	Configuration	m	$M = m^N$	T
3	uud	2.73	20.39	.77
4	uuud	2.91	72.66	.65
	duud	2.48	37.90	.78
	udud	1.54	5.67	1.00
	uudd	1.17	1.87	.94
5	uuuud	2.97	232.82	.56
	duuud	2.85	190.92	.60
	uduud	2.58	116.39	.69
	uudud	1.83	20.54	.95
	uuudd	1.35	4.61	.99
	dduud	1.23	2.81	.95
6	uuuuud	2.99	716.91	.50
	duuuud	2.95	667.83	.51
	uduuud	2.88	571.43	.53
	dduuud	2.83	519.66	.55
	uuquud	2.62	326.78	.63
	duduud	2.55	274.97	.66
	udduud	2.40	191.94	.72
	uuudud	1.92	51.04	.91
	duudud	1.68	23.04	.98
	ddduud	1.53	12.84	1.00
	uuuudd	1.44	9.22	1.00
	ududud	1.38	7.15	.99
	duuudd	1.21	3.15	.94
	uuddud	1.16	2.48	.91
	uduudd	1.13	2.17	.89
	uududd	1.09	1.70	.86
	uuuddd	1.06	1.47	.84

Off-the-shelf prism beam expanders

Optimal solutions are useful when appropriate resources are available for the purchase of custom prisms: invariably the optimal prism apex angles take on somewhat obscure values. At the same time, however, the prism apex angles are the least sensitive parameters in the optimization.⁴ Consequently, approximately optimal solutions are obtainable with off-the-shelf prisms, provided that the appropriate prism configuration is employed. To illustrate this last point and, more importantly, to provide useful results for the casual or price-conscious prism-beam-expander builder, we present the results of a numerical optimization of prism beam expanders employing only 45-45-90 prisms of the same material with refractive index 1.5. The optimization technique was a Newton-Marquardt routine employing a Cholesky decomposition subroutine and penalty functions to incorporate the magnification and dispersion constraints. The use of numerous initial guesses served as a check that the results obtained represented global, rather than local, optima. Results containing external prism exit angles of greater than 45 degrees were assumed to violate the assumption of broad-band AR-coated exit faces⁵ and were therefore rejected and not plotted in Figures 3-6. Such rejection of results only proved necessary for extremely low ($M < 3$) magnifications and hence was relatively unimportant.

Figure 3 plots the optimal transmission vs. magnification for several two-, three-, and four-prism single-material beam expanders. Clearly the only achromatic two-prism configuration is the up-down arrangement. What is not so clear, however, is that the only useful three-prism configuration is up-up-down: the other two (up-down-down and down-up-down) do not generally yield achromatic results with 45-degree apex angle prisms. Of the seven possible four-prism designs, we plot only the two best configurations, up-up-up-down and down-up-up-down, and a seemingly useful configuration, up-down-up-down, for comparison. The poor performance of the up-down-up-down configuration dispels the notion that achromatic prism beam expanders are best constructed from achromatic pairs of prisms. Such a notion merely adds another constraint to an already quite constrained problem and is clearly

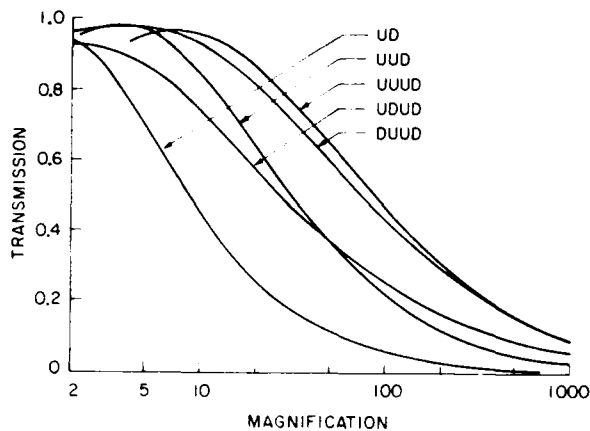


Figure 3. Optimal transmission vs. magnification for several two-, three-, and four-prism achromatic prism beam expanders employing 45° apex angle prisms.

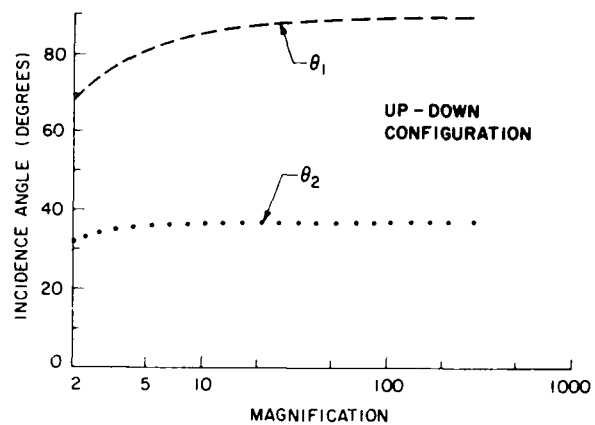


Figure 4. Incidence angles for optimal achromatic two-prism beam expanders using 45° apex angle prisms (up-down configuration).

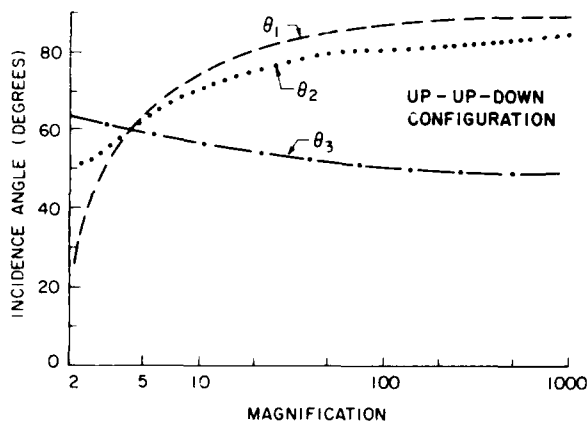


Figure 5. Incidence angles for optimal achromatic three-prism beam expanders using 45° apex angle prisms (up-up-down configuration).

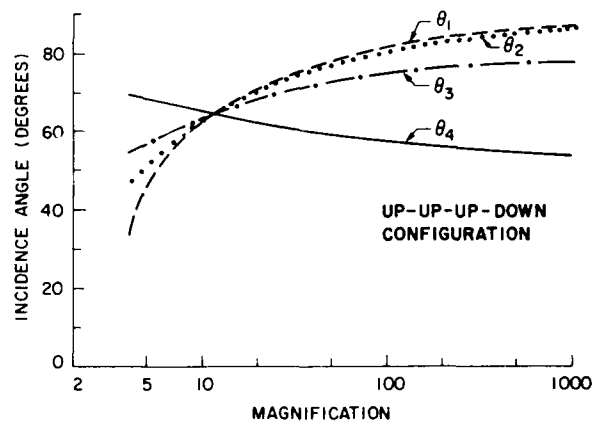


Figure 6. Incidence angles for optimal achromatic four-prism beam expanders using 45° apex angle prisms (up-up-up-down configuration).

counterproductive. Finally, we point out that, when compactness is important or the number of prisms must be limited for some other reason, the three-prism up-up-down design⁵ offers an attractive, if somewhat unintuitive, solution to the problem of an intermediate-magnification prism beam expander. Observe that it yields better transmission than the four-prism up-down-up-down configuration for magnifications up to about 60.

Figures 4, 5 and 6 give the optimal incidence angles for use in several of these "off-the-shelf" prism beam expanders. For two-prism designs, we see that the incidence angle of the second prism remains approximately constant and relatively small, while that of the first prism approaches 90 degrees for large magnifications. Consequently, most of the magnification occurs at the first prism as does most of the reflection loss. Interestingly, because the second prism's incidence angle remains significantly below Brewster's angle, some additional loss occurs at the second prism's entrance face. The use of a smaller apex angle in the second prism avoids this problem and achieves superior performance,⁷ but, in general, two-prism achromatic beam expanders are a poor choice for magnifications in excess of relatively small values of about 4.

The three-prism, up-up-down configuration is generally a better choice for magnifications in the 3 to 10 range, and in some specialized applications for larger magnifications, also.⁴ Similarly, the four-prism up-up-up-down configuration⁵ is probably the best choice for higher magnifications.

Notice that in all expanders employing only 45 degrees apex angle prisms of the same material, the final prism's incidence angle decreases with increasing magnification, while the other prism incidence angles increase (as expected). This decrease in the final incidence angle occurs because, as the magnification increases, the final prism's dispersion tends to dominate.⁶ The optimization process thus forces smaller and smaller incidence angles on the final prism to decrease its magnification because the dispersions of all the other prisms are divided by this magnification in the equation for the total prism beam expander dispersion.⁷ A good design then—that is, one that employs close to optimal apex angles—will not require widely differing incidence angles for the first $N-1$ and last prisms. The use of only 45 degree apex angle prisms is thus not in general the best way to obtain nearly optimal achromatic prism beam expanders; just the same, the solutions given here are probably only a few per cent away from optimality over most of the plotted range. Finally, observe that equality of all of the incidence angles occurs approximately where true optimality (given in Table 1) occurs.

Conclusions

On the whole, the problem in achromatic-beam-expander design is to compensate for the otherwise dominating dispersion of the final prism. Several methods of attack are available: (1) the use of the up-up-...-up-down configuration; (2) the use of low-dispersion material for the final prism; (3) the use of a smaller incidence angle for the final prism; and finally, (4) the use of a smaller apex angle for the final prism. None of these solutions is ideal in terms of optimality, convenience, simultaneous achromaticity and thermal stability, cost, and simple and quick assembly from currently available off-the-shelf prisms; the particular solution chosen must necessarily depend on the specific application.

Acknowledgements

The authors would like to thank the Air Force Office of Scientific Research for the financial support that made this research possible.

References

1. Klauminzer, G. K., "New High-Performance Short-Cavity Dye-Laser Design," IEEE. J. Quant. Electron., Vol. QE-13, p. 92D. 1977.
2. Duarte, F. J., "Prism-Grating Systems for Laser Wavelength Measurements," J. Phys. E: Sci. Instrum., Vol. 16, p. 599. 1983.
3. Fejer, M. M., G. A. Magel, and R. L. Byer, "High-Speed, High-Resolution Fiber-Diameter-Measurement System," Appl. Opt., to be published.
4. Trebino, R., "Achromatic N-Prism Beam Expanders: Optimal Configurations," Appl. Opt., Vol. 24, Apr. 15, 1985.
5. Patent pending.
6. Corion Corporation, catalog.
7. Barr, J. R. M., "Achromatic Prism Beam Expanders," Opt. Commun., Vol. 51, p. 41. 1984.

Rise-fall ambiguities and their removal from frequency-domain nonlinear-optical techniques

Rick Trebino

Combustion Research Facility, Sandia National Laboratories, Livermore, California 94550

Charles E. Barker and Adnah G. Kostenbauder

E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305

Received April 21, 1988; accepted June 22, 1988

Most frequency-domain nonlinear-optical techniques measure only the magnitude of the Fourier transform of a temporal response and, hence, do not uniquely determine the response. We show that, for the commonly used response, $h(t) = A \exp(-t/\tau_f) + B \exp(-t/\tau_s)$, where $A \equiv \alpha/\tau_f$ and $B \equiv (1 - \alpha)/\tau_s$, the spectral line shape can always be fitted by two different values of α . A measurement of the optical Kerr transient of carbon disulfide illustrates this ambiguity. We also demonstrate a single-scan method that is free from such ambiguities. It involves adding coherent background with a nonzero quadrature-phase component, obtained simply by proper choice of probe wavelength.

In principle, each function of a Fourier-transform pair uniquely determines the other. In practice, however, nonlinear-optical techniques that operate in the frequency domain often yield only the magnitude, and not the phase, of the frequency response.¹⁻⁵ As a result, the corresponding temporal response is not uniquely determined. Obvious (and trivial) ambiguities include a constant phase factor, displacement of the origin, and simultaneous inversion and complex conjugation.⁶ Less obvious, and far less trivial, ambiguities also exist, even if the class of possible temporal responses is restricted.⁶⁻⁸ The study of such ambiguities, known as the phase-retrieval problem, has received much attention.⁸

In this Letter we consider techniques that measure the magnitude of the frequency response and study a specific, commonly used temporal response,^{1,9} the sum of two exponentials:

$$h(t) = \theta(t)[A \exp(-t/\tau_f) + B \exp(-t/\tau_s)], \quad (1)$$

where $\theta(t)$ is the unit step function, τ_f and τ_s are fast and slow time scales, respectively, $A \equiv \alpha/\tau_f$, $B \equiv (1 - \alpha)/\tau_s$, and α is the relative weight. All parameters are real.^{1,9} We show that knowledge of only the magnitude of the frequency response of this decay is not sufficient to permit us to specify uniquely the parameter α . Specifically, for every measurable line shape there are always two possible values of α . We refer to this ambiguity as a rise-fall ambiguity because it often results in a temporal response involving a rapid fall and then a slower fall having an identical line shape to that of a different temporal response, one involving a rapid rise and then a slower fall. This type of ambiguity could be particularly troublesome, especially when each of the above responses corresponds to a different physical phenomenon or when rises and falls both potentially occur.¹⁰⁻¹²

We demonstrate this ambiguity in experiments that use a nonlinear-optical technique¹ that involves mea-

suring the strength of an induced grating as a function of the frequency difference between the two writing beams, $\omega_1 - \omega_2$. A third beam, at frequency ω_3 , probes the grating, producing a signal beam at the frequency $\omega_0 = (\omega_1 - \omega_2) + \omega_3$. Measuring the diffraction efficiency as ω_1 is scanned yields the magnitude of the Fourier transform of the material temporal response. This technique is one of a class of frequency-domain nonlinear-optical methods,¹⁻⁵ which generally measure the magnitude of the Fourier transform of the temporal response of a material and which have been used to study a wide range of effects. Despite the wide use of these methods, however, potential ambiguities have not to our knowledge been recognized.

In other contexts, many techniques have been suggested for avoiding or removing such ambiguities. They include complementary aperturing,¹³ exponential filtering,¹⁴ using multiple scans under slightly different conditions,¹⁵ and introducing background from a point source separate from the original function.¹⁶ Burge *et al.*¹⁷ have suggested adding a sufficient amount of constant, coherent background of arbitrary phase to remove all zeros and then Hilbert transforming the natural logarithm of the modulus of the result. Similarly, one could use optical homodyne detection,¹⁸ in which a large amount of coherent in-phase or quadrature-phase background is added to the signal beam, yielding the real or imaginary component of the line shape. Two scans, one with each phase, yield the complete Fourier transform. Alternatively, with one component obtained from one scan, a Hilbert transform generates the other.

None of these methods appears practical for general nonlinear-optical spectroscopy, however. Consequently, we introduce and demonstrate a new method, which also involves the introduction of a constant, coherent background. In this method, the background must necessarily contain some quadrature phase (with respect to that of the signal at line center).¹⁹ It is to be distinguished from other background

schemes (such as that of Ref. 17), in which a minimum background *magnitude* is specified. It should also be distinguished from optical homodyne detection, which requires a specific phase and a large amount of background. Both of the last two methods are difficult to implement in nonlinear-optical techniques.

The scheme that we present is quite simple. It is not material specific, and it requires only a single scan. It involves tuning the probe frequency to a value $\sim 1/\tau$ or $\sim -1/\tau$ from that of the fixed-frequency excitation beam, where τ is an effective time scale of the material response. This introduces into the signal beam non-negligible contributions from a new process¹: $(\omega_3 - \omega_2) + \omega_1$. In this process, the beams at frequencies ω_3 and ω_2 write the grating, and the (variable-frequency) beam at ω_1 probes it. Because the frequency difference between the two writing beams is constant, the strength of this contribution is constant throughout a scan (the probe transition is assumed to be broad compared with the scan range). This process has the same phase-matching properties, coherence properties, and strength as does the $(\omega_1 - \omega_2) + \omega_3$ process. Finally, it has a strong component of quadrature phase. The main advantage of this method in nonlinear-optical techniques is that the $(\omega_3 - \omega_2) + \omega_1$ background automatically has constant phase with respect to the signal, independent of the phase fluctuations of the input beams. In contrast, maintaining constant phase in optical-homodyne experiments is usually quite difficult.

We now demonstrate these statements. The measured line shape is the squared magnitude of the sum of the Fourier transform (with respect to $\omega \equiv \omega_1 - \omega_2$) of Eq. (1) and some coherent background, $\beta + i\gamma$:

$$|\Gamma_{\alpha\beta\gamma}(\omega)|^2 = \left| \frac{1 + i\omega\tau_a}{(1 + i\omega\tau_f)(1 + i\omega\tau_s)} + \beta + i\gamma \right|^2, \quad (2)$$

where $\tau_a \equiv \alpha\tau_s + (1 - \alpha)\tau_f$. We ask whether there exist α' , β' , and γ' such that $|\Gamma_{\alpha'\beta'\gamma'}(\omega)|^2 = |\Gamma_{\alpha\beta\gamma}(\omega)|^2$ for all ω . We find that this is the case, and we obtain two solutions: a trivial one ($\alpha' = \alpha$, $\beta' = \beta$, and $\gamma' = \gamma$) and a nontrivial one:

$$\alpha' = -\frac{2\tau_f}{\tau_s - \tau_f} - 2(\beta - i\gamma)\frac{\tau_s + \tau_f}{\tau_s - \tau_f} - \alpha, \quad (3)$$

with $\beta' = \beta$ and $\gamma' = -\gamma$. The nontrivial solution represents the troublesome ambiguity, and the expression for it in Eq. (3) is exact. Equation (3) can also be obtained by writing $\Gamma_{\alpha\beta\gamma}(\omega)$ as a single quotient and taking the complex conjugate of the numerator only. This produces a nontrivial change in the response but does not change its magnitude.²⁰

The introduction of quadrature-phase coherent background, $i\gamma$, removes the ambiguity because it forces the value of α' off the real line. Since a complex relative weight violates the assumption of a real decay, it is no longer physically acceptable. Even if complex weights are allowed, quadrature phase may remove the ambiguity because the alternative solution for α is necessarily accompanied by ambiguity in γ : $\gamma' = -\gamma$. Thus, if only its sign is known, the presence of coherent, quadrature-phase background removes rise-fall

ambiguities. Note that any nonzero amount of quadrature-phase background is sufficient to remove the ambiguity, although, in practice, experimental noise implies a minimum value.

We have observed this ambiguity in frequency-domain experiments to measure the optical Kerr response in carbon disulfide (the technique and experimental apparatus are discussed in Ref. 1). This response has been measured^{10,13} and fitted with sum-of-two-exponentials decays. The resulting ranges of fit parameters in these measurements are $\tau_f = 100$ –360 fsec, $\tau_s = 1.3$ –2.2 psec, and $\alpha = 0.16$ –0.51. We neglect other time scales (e.g., the electronic component and the recently observed ultrafast rise^{10,12}).

To fit the data, we use an eight-parameter model involving the above parameters and, in addition, a baseline, a scale parameter, and a (small) wavelength shift parameter. Figure 1(a) shows a previously reported experimental spectrum for carbon disulfide in the absence of quadrature-phase background.^{10,21} In this experiment, $\omega_3 - \omega_2 \approx 3.8 \times 10^{13} \text{ sec}^{-1}$ (200 cm^{-1}), which is much greater than $1/\tau$ ($\approx 10^{12} \text{ sec}^{-1}$), so that the $(\omega_3 - \omega_2) + \omega_1$ process contributes negligible coherent background to the signal beam. Fitting these data, we obtain two solutions, which together compose the solid line in Fig. 1(a). All the parameters of these two fits are nearly identical, except the relative-weight parameters, α , which differ by a large amount: $\tau_s = 1.32$ psec, $\tau_f = 0.19$ psec, $\beta = 0.12$, $\gamma = 0.001$, and $\alpha =$

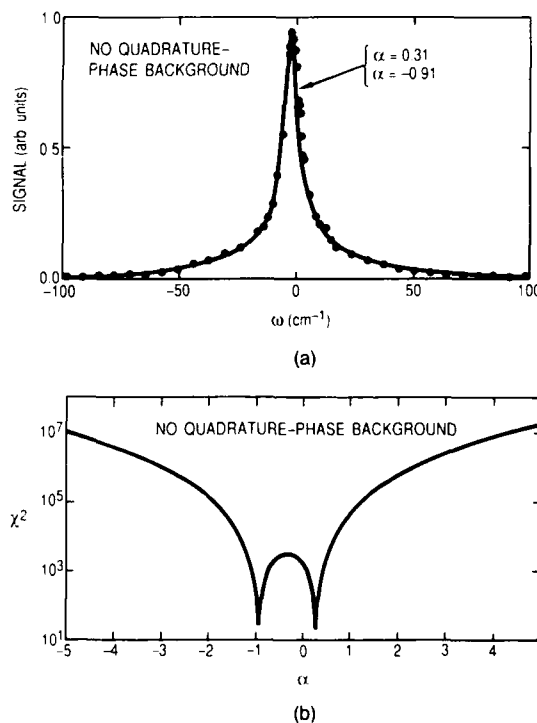


Fig. 1. Experimental results with negligible quadrature-phase coherent background. (a) Diffraction efficiency versus frequency difference, $\omega = \omega_1 - \omega_2$. The solid line indicates the best fit, obtained for two different values of α : -0.91 and $+0.31$. (b) Plot of $\chi^2(\alpha)$ versus α for the data in (a). Observe the presence of two minima with about the same minimum value, indicating ambiguity in the fit for the parameter α .

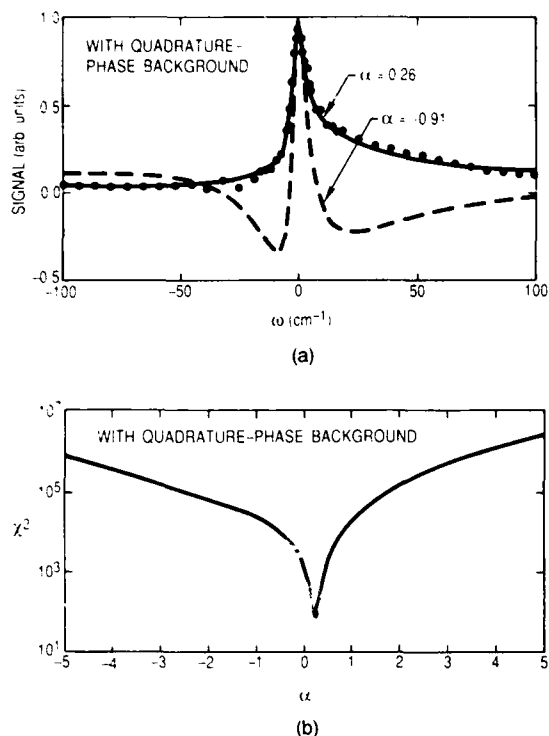


Fig. 2. Experimental results with substantial quadrature-phase coherent background. (a) Diffraction efficiency versus frequency difference, $\omega = \omega_1 - \omega_2$. The solid line indicates the best fit, obtained for a single value of α of +0.26. The dashed line indicates a fit using $\alpha = -0.91$. (A negative baseline has been added to both curves to force their asymptotes to zero.) (b) Plot of $\chi^2(\alpha)$ versus α for the data in (a). Observe the presence of only one minimum, indicating that ambiguity in the fit for the parameter α has been removed.

-0.91 or +0.31. That these two solutions represent the predicted ambiguity is demonstrated in Fig. 1(b), in which $\chi^2(\alpha)$ is the value of the normalized sum of the squares of the weighted residuals for a given value of α , with all other parameters fixed at their optimum values. Observe that there are two very steep minima, representing the two possible solutions, which satisfy Eq. (3) to within a few percent.

We also performed an experiment using a probe wavelength closer to that of the fixed-frequency excitation beam: $\omega_3 - \omega_2 = 4.6 \times 10^{12} \text{ sec}^{-1}$ (25 cm⁻¹). Because $|\omega_3 - \omega_2| \sim 1/\tau_f$ in this experiment, a much greater amount of coherent background with a substantial quadrature-phase component contributes, yielding a significant amount of asymmetry. Figure 2(a) shows the data and two theoretical curves. The solid curve, which fits the data well, is the result of a curve-fitting run that optimized all parameters, yielding $\tau_s = 1.42 \text{ psec}$, $\tau_f = 0.19 \text{ psec}$, $\beta = 1.40$, $\gamma = -1.04$, and $\alpha = +0.26$. The dashed curve, which does not fit the data well, results from the use of all these parameters except α , for which we use the negative solution obtained previously (-0.91). Allowing all the parameters to float does not improve the fit for any negative value for α . In addition, the values obtained for the time scales in this experiment in the positive- α fit are

nearly identical to those obtained in the positive- α solution in the experiment without coherent background. To verify that no other solution exists for α , we plot in Fig. 2(b) the function $\chi^2(\alpha)$ for these data, showing that the only minimum occurs at the correct positive- α location. Thus we confirm that the introduction of quadrature-phase coherent background removes the ambiguity in α .

In this Letter, we have limited the discussion to zero-frequency resonances, but the results are relevant to other coherent frequency-domain methods, such as Raman spectroscopy. A future publication will discuss more-general line shapes and will give a simple procedure for extracting real responses from Fourier-transform-magnitude data in the presence of quadrature-phase background without the need to assume a theoretical model.

This research was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Chemical Sciences Division, and the U.S. Air Force Office of Scientific Research.

References

1. R. Trebino, C. E. Barker, and A. E. Siegman, *IEEE J. Quantum Electron.* **QE-22**, 1413 (1986).
2. L. J. Rothberg and N. Bloembergen, *Phys. Rev. A* **30**, 2327 (1984).
3. J. J. Song, J. H. Lee, and M. D. Levenson, *Phys. Rev. A* **17**, 1439 (1978).
4. T. Yajima, H. Souma, and Y. Ishida, *Phys. Rev. A* **17**, 324 (1978).
5. F. Keilmann, *Appl. Phys.* **14**, 29 (1977).
6. J. R. Fienup, *J. Opt. Soc. Am.* **4**, 118 (1987).
7. E. J. Akutowicz, *Trans. Am. Math. Soc.* **83**, 179 (1956).
8. H. Stark, ed., *Image Recovery: Theory and Applications* (Academic, Orlando, Fla., 1987).
9. W. Yu, F. Pellegrino, M. Grant, and R. R. Alfano, *J. Chem. Phys.* **64**, 2648 (1977).
10. C. Kalpouzos, W. T. Lotshaw, D. McMorrow, and G. A. Kenney-Wallace, *J. Phys. Chem.* **91**, 2028 (1987).
11. J.-M. Halbout and C. L. Tang, *Appl. Phys. Lett.* **40**, 765 (1982).
12. S. Ruhman, L. R. Williams, A. G. Joly, B. Kohler, and K. A. Nelson, *J. Phys. Chem.* **91**, 2237 (1987).
13. D. L. Misell, R. E. Burge, and A. H. Greenaway, *J. Phys. D* **7**, L27 (1974).
14. C. L. Mehta, *Nuovo Cimento* **36**, 202 (1965).
15. D. L. Misell, *J. Phys. D* **6**, 2200 (1963).
16. C. L. Mehta, *J. Opt. Soc. Am.* **58**, 1233 (1968).
17. R. E. Burge, M. A. Fiddy, A. H. Greenaway, and G. Ross, *J. Phys. D* **7**, 61 (1974).
18. M. D. Levenson and G. L. Eesley, *Appl. Phys.* **19**, 1 (1979).
19. Ironically, the two frequency-domain methods^{3,5} that do involve coherent background contain in-phase, but not quadrature, background.
20. The existence of this ambiguity and its removal by the addition of quadrature-phase coherent background can also be seen when using Blaschke products.⁷ Ambiguity removal also follows from the observation that the Fourier-transform magnitude of a real decay is even, and adding $i\gamma$ breaks this symmetry.
21. In the course of this work, the fit to the data of this experiment was slightly improved, yielding the parameter values reported herein.

100-31

Frequency-Domain Observation of the Ultrafast Inertial Response of the Optical Kerr Effect in CS₂

Charles E. Barker
E. L. Ginzton Laboratory
Stanford University
Stanford, California 94305
(415) 723-0277

Rick Trebino
Combustion Research Facility
Sandia National Laboratories
Livermore, California 94550
(415) 294-2893

A. G. Kostenbauder and A. E. Siegman
E. L. Ginzton Laboratory
Stanford University
Stanford, California 94305
(415) 723-0277

Abstract

Time domain experiments using femtosecond pulses have recently revealed the inertial aspect of the nuclear orientation component of the optical Kerr effect in carbon disulfide. We present frequency domain measurements, performed with the tunable-laser-induced grating technique, that also demonstrate the need to incorporate inertial behavior in models of the nuclear orientation contribution to this ultrafast response.

Frequency-Domain Observation of the Ultrafast Inertial Response of the Optical Kerr Effect in CS₂

Charles E. Barker
E. L. Ginzton Laboratory
Stanford University
Stanford, California 94305
(415) 723-0277

Rick Trebino
Combustion Research Facility
Sandia National Laboratories
Livermore, California 94550
(415) 294-2893

A. G. Kostenbauder and A. E. Siegman
E. L. Ginzton Laboratory
Stanford University
Stanford, California 94305
(415) 723-0277

Introduction

Since it was first observed in 1964, the optical Kerr effect,^{1,2} particularly as it is displayed by carbon disulfide, has been the subject of much study by a variety of laser techniques.³⁻¹⁰ Measurements of the temporal response of the optical Kerr effect are not only important for the design of ultrafast optical devices,¹¹ but also provide a valuable probe of the underlying molecular dynamics.¹²⁻¹⁴ While light scattering experiments³ and time domain experiments with subpicosecond pulses^{7,8} have shown that the temporal response of CS₂ is not adequately described by a simple orientational diffusion model, the recent time domain experiments by Kenney-Wallace and co-workers^{15,16} and by Nelson and co-workers^{17,18} are the first measurements to provide clear evidence of the inertial nature of the nuclear orientation component of the optical Kerr effect.

Although the resolution of time domain techniques continues to improve with advances in the generation of ultrashort laser pulses, in principle the same information, with equal or better resolution, is available in the frequency domain. The tunable-laser-induced grating (TLIG) technique is one such frequency domain method possessing femtosecond resolution.¹⁹ A tunable-laser-induced grating experiment measures the diffraction efficiency of a grating induced in the complex index of refraction of a sample material by the interference of a pair of excitation beams whose frequency difference can be varied. When this frequency difference is small, the interference pattern moves slowly and the material is able to follow it, producing maximum diffraction efficiency. On the other hand, when the frequency difference is large, the interference pattern moves so rapidly that the material is unable to follow and the diffraction efficiency is, consequently, very small. It can be shown that the diffraction efficiency as a function of frequency difference between the excitation beams is the square magnitude of the Fourier transform of the system impulse response; in four wave mixing terminology, the diffraction efficiency is proportional to the square magnitude of the third order nonlinear optical susceptibility.¹⁹ Among the advantages of the tunable-laser-induced grating technique are the capability to measure a single element of the nonlinear susceptibility tensor and to choose a phase matching geometry which provides sufficient bandwidth to study extremely fast events. The various aspects of this technique have been discussed in detail elsewhere.¹⁹⁻²¹ Using the tunable-laser-induced grating technique, we have measured the ultrafast response of the optical Kerr effect in CS₂, and we report frequency domain measurements which yield evidence of the inertial character of the optical Kerr response of CS₂.

Experimental Technique

The apparatus we used to perform tunable-laser-induced grating experiments is shown in Figure 1. It consists essentially of three pulsed dye lasers, all pumped by the same frequency-doubled Nd:YAG laser.¹⁹ The YAG laser is a Q-switched, unstable-resonator laser which provides approximately 110 mJ, 10 ns pulses of 1.064 μm radiation at a repetition rate of 10 Hz. This light is frequency-doubled in a KD*P crystal to produce 18-21 mJ of 532 nm radiation which is split into three equal-energy beams to pump three dye lasers. Each dye laser is of a prism-beam-expander/grazing-incidence hybrid design and provides 0.2-0.7 mJ, 7 ns pulses with spectral line widths of 0.1-0.2 cm^{-1} using rhodamine 6G dye.

After being collimated, each dye-laser beam is focussed by 15-35 mm focal-length lenses into a 1 mm thick glass sample cell; broadband polarizers are placed in the beam paths adjacent to the cell to ensure that the beams entering the sample are highly polarized. The two excitation beams are counterpropagating, and the probe beam is incident at an angle of a few degrees with respect to the fixed frequency excitation beam. The diffracted signal beam is passed through a broadband polarizer, detected with a photodiode, and amplified for measurement by the laboratory computer. The diffraction efficiency curve is obtained by scanning the wavelength of one of the excitation lasers.

The measured diffraction efficiency curves are curve-fit to a lineshape function of the general form:

$$\eta(\Delta\omega_{12}) = C_0 + H \left[\left(R(\Delta\omega_{12}) + C_R \right)^2 + \left(I(\Delta\omega_{12}) + C_I \right)^2 \right], \quad (1)$$

where C_0 is the incoherent background, H is the overall peak height above this background, C_R and C_I are the real and imaginary parts, respectively, of the coherent background, and $R(\Delta\omega_{12})$ and $I(\Delta\omega_{12})$ are the real and imaginary parts, respectively, of

the frequency response function $\Gamma(\Delta\omega_{12})$ which is the Fourier transform of the material's impulse response function.¹⁹ A simplex algorithm^{22,23} is used to minimize the sum of the weighted least squares, χ^2 . The residuals of the fit are subjected to a variety of goodness-of-fit tests, and fits to models with different numbers of parameters are compared using the F-test.²⁴

Results

The data from tunable-laser-induced grating measurements of the optical Kerr effect in CS₂ at 298 ± 1 K are shown in Figures 2-7 along with fits to different models and plots of the corresponding time-domain impulse response functions; the data presented in all six figures are from the same scan. A set of four such scans of the nonlinear susceptibility tensor element $\chi^{(3)}_{xyyx}$ were obtained with orthogonally polarized excitation beams whose wavelengths were centered at 575 nm and a probe beam whose wavelength was 572 nm. Each scan consists of 57 data points, each of which is an average over 600 laser shots. The use of orthogonally polarized excitation beams allows us to avoid the spectral "contamination" of intensity-grating effects which obscure the optical Kerr spectrum,^{19,25,26} and, as a further benefit, allows probe and signal polarizations which greatly reduce scattered light at the signal detector. The effective temporal resolution of these scans is better than 40 femtoseconds, which is the exponential decay time corresponding to the frequency scan range. The phase matching bandwidth, determined by the excitation and probe beam geometry, is approximately 500 cm^{-1} (full width) for these experiments; although this factor is neglected in the curve fitting, its primary effect is a slight increase in the observed values of the time constants obtained by curve fitting. The slight asymmetry in the curves is due to the presence of a small amount of coherent background,¹⁹ which is included in the fit.

$$\Delta\omega = 5 \text{ fs cm}^{-1}$$

$$\Delta\omega = \frac{2}{\Delta t}$$

Under the assumption that the Born-Oppenheimer approximation is valid, mathematical models of the dynamical response of the optical Kerr effect can be expressed as the sum of an electronic contribution and a nuclear contribution.²⁷ Because it is viewed as instantaneous even at optical frequencies, the nonresonant electronic response appears in Tunable-Laser-Induced Grating lineshapes as part of the coherent background.²⁶ Thus the spectral peak obtained in a Tunable-Laser-Induced Grating measurement of the response of an optical Kerr liquid is due to the noninstantaneous motion of the nuclei which comprise the molecules of the liquid. Several models of the dynamical response of this nuclear contribution to the optical Kerr effect in CS₂ have been proposed.^{4,7,8,13,15-18,28-30} We present below the results of fitting six of these models to our data.

Early treatments of the dynamical response of the nuclear contribution to the induced birefringence $\Delta n(t)$ were based on a simple orientational diffusion model,^{31,32} and led to a first order linear differential equation for the temporal behavior of the transient birefringence induced by a picosecond light pulse.²⁸ The impulse response of this first order differential equation is given by:

$$\Delta n(t) = A \exp(-t/\tau_{or}), t \geq 0, \quad (2)$$

which gives a frequency response that is a single lorentzian whose width is determined by the orientational diffusion lifetime τ_{or} :

$$\Gamma(\omega) = \frac{A\tau_{or}}{1 + i\omega\tau_{or}}. \quad (3)$$

The inability of this model to explain the high frequency wings observed in light scattering spectra^{3,13,33} and the apparently non-single-exponential decays measured in subpicosecond time-domain experiments^{7,8} led researchers to model the nuclear or orientational response of CS₂ as the sum of two or more components occurring on different timescales:³⁴

$$\Delta n(t) = \sum_i \Delta n^{(i)}(t). \quad (4)$$

Assuming that the $\Delta n^{(i)}$ are given by Eq. (2), one of the simplest response functions displaying this form is a two exponential decay:

$$\Delta n_1(t) = A \exp(-t/\tau_1) + (1-A)\exp(-t/\tau_2), \quad t \geq 0, \quad (5)$$

which has the Fourier transform:

$$\Gamma_1(\Delta\omega) = \frac{A\tau_1}{1 + i\Delta\omega\tau_1} + \frac{(1-A)\tau_2}{1 + i\Delta\omega\tau_2}. \quad (6)$$

Figure 2(a) shows the Tunable-Laser-Induced Grating data fit to the square magnitude of this frequency response function as prescribed by Eq. (1). Fitting this model to a set of four separate scans produced the following average values for parameters: $A = 0.36 \pm 0.03$, $\tau_1 = 1.21 \pm 0.06$ ps, and $\tau_2 = 0.182 \pm 0.020$ ps, with χ^2 values for the individual fits ranging from 13.55 to 18.99. Although these parameter values are in rough agreement with the subpicosecond time-domain results of Halbout and Tang⁷ and Greene and Farrow⁸ as shown in both Table 1 and Figure 2(b), both goodness-of-fit tests applied to the residuals of the fits and a visual examination of the fit shown in Figure 2(a) (in the region around 50 cm^{-1}) suggest that this model does not provide an adequate description of the data. It does, however, indicate approximate timescales for the optical Kerr response.

The recent time-domain experiments of Kenney-Wallace and co-workers^{15,16} and Nelson and co-workers^{17,18} indicate that the temporal response of the nuclear contribution to the optical Kerr effect exhibits an inertial character, i.e. a finite risetime. An inertial model has also been invoked to explain the wings of light scattering spectra²⁹ and self-trapping phenomena in CS_2 .⁴ The simplest means by which inertial behavior can be incorporated into this framework is to model CS_2 molecules as orientational damped harmonic oscillators.^{4,16,29} This leads to a second order linear differential equation similar to the ones proposed to explain filament spectra⁴ and transmission through an optical Kerr

gate.¹⁶ Assuming the initial condition $\Delta n^{(i)}(t = 0) = 0$, the impulse response of a second order linear ordinary differential equation is given in the overdamped case by:

$$\Delta n^{(i)}(t) = A_i \exp(-t/\tau_i) \sinh(t/\beta_i) , \quad (7)$$

and in the underdamped case by:

$$\Delta n^{(i)}(t) = A_i \exp(-t/\tau_i) \sin(t/\beta_i) . \quad (8)$$

Several overall nuclear impulse responses have been constructed from the functions given in Eqs.(2), (7), and (8) and applied to the time domain results of Kenney-Wallace *et al*^{15,16} and Nelson *et al.*^{17,18} Modeling the faster of two components of the nuclear response as an overdamped oscillator, Nelson and co-workers¹⁷ have proposed the following response function:

$$\Delta n_2(t) = \exp(-t/\tau_1) + B(\exp(-t/\tau_2) - \exp(-t/\tau_3)), \quad t \geq 0, \quad (9)$$

which has a frequency response given by:

$$\Gamma_2(\Delta\omega) = \frac{\tau_1}{1 + i\Delta\omega\tau_1} + \frac{B\tau_2}{1 + i\Delta\omega\tau_2} - \frac{B\tau_3}{1 + i\Delta\omega\tau_3} . \quad (10)$$

Note that the overall response described by Eq.(9) starts from a nonzero value at $t = 0$.

Figure 3(a) shows our data fit to the square magnitude of the frequency response given by Eq. (10). Again, averaging over the best-fit parameter values for the four spectral scans we obtain the following values: $B = 58.9 \pm 21.5$, $\tau_1 = 1.24 \pm 0.04$ ps, $\tau_2 = 0.134 \pm 0.010$ ps, and $\tau_3 = 0.123 \pm 0.010$ ps., where the error values stated for the parameters τ_2 and τ_3 have been increased to more accurately reflect experimental noise and uncertainties. A plot of the impulse response given by Eq. (9), normalized for unit height and using the average values we observe for the parameters, is displayed in Figure 3(b). Figure 3(b) and Table 2 show that these values agree quite well with those reported by Nelson and co-workers.¹⁷ The

large variation in the value of the relative weight parameter B, both in our curve-fitting results from scan-to-scan as reflected in the large error quoted for this parameter and in comparison with the value reported by Nelson *et al*, seems to be compensated by much smaller variations in the lifetime parameters τ_i ; hence these variations appear to be due to the well-known difficulties associated with the fitting of multiple exponentials.²² The χ^2 values for our fits range from 4.74 to 7.89. Of course, whenever the number of degrees of freedom in a fit are lowered by the introduction of additional parameters into the model, the χ^2 of the fit will decrease. Consequently, the significance of the decrease in χ^2 must be determined by the use of an F-test.²⁴ For each scan, the F-test indicated that the more complicated model described by Eqs. (9) and (10) was justified at a 99.9% confidence level—in other words, the probability of such a large decrease in χ^2 occurring by random chance was less than 0.1%.

Further improvement in the goodness-of-fit can be obtained by rearranging the relative weight parameters in Eq. (9) so that the impulse response starts from zero, i.e. $\Delta n(t=0) = 0$. Such a model, which is similar to those used by Kenney-Wallace and co-workers¹⁶ and which can be obtained without the addition of more parameters, can be written in the following way:

$$\Delta n_3(t) = A \exp(-t/\tau_1) + (1-A) \exp(-t/\tau_2) - \exp(-t/\tau_3), \quad t \geq 0. \quad (11)$$

The Fourier transform of this response is given by:

$$\Gamma_3(\Delta\omega) = \frac{A\tau_1}{1+i\Delta\omega\tau_1} + \frac{(1-A)\tau_2}{1+i\Delta\omega\tau_2} - \frac{\tau_3}{1+i\Delta\omega\tau_3}. \quad (12)$$

The results of fitting this model to our Tunable-Laser-Induced Grating data are shown in Figure 4(a), and a plot of the impulse response function given by Eq. (11), using the average values obtained for the parameters and normalized for unit height, is displayed in

Figure 4(b). The average values of the best-fit parameters for the set of four scans are: $A = 0.009 \pm 0.003$, $\tau_1 = 1.19 \pm 0.04$ ps, $\tau_2 = 0.112 \pm .010$ ps, and $\tau_3 = 0.106 \pm 0.010$ ps, where we have again increased the uncertainties stated for τ_2 and τ_3 by a factor of 4 in order to better represent experimental error. If we define an "effective" B for Eq. (11) as $B = (1-A)/A$, Table 2 shows that these values also agree well with those reported by Nelson and co-workers.¹⁷ The χ^2 values for these fits range from 3.41 to 6.09; on average this is approximately a 27% decrease over the χ^2 values for the fits to the model described by Eq. (9), which has the same number of parameters. A close examination of Figures 3(b) and 4(b) reveals that the only significant difference between these two responses is the behavior very near $t = 0$. Consequently, it appears that the contribution of the nuclear orientation to the optical Kerr effect is zero at $t = 0$.

Let us now turn to the case of an underdamped response for the fast component of the orientational response. For this case, Nelson and co-workers¹⁸ have presented the following response function:

$$\Delta n_4(t) = \exp(-t/\tau_1) + C \exp(-t/\tau_2) \sin(t/\beta), \quad t \geq 0, \quad (13)$$

as a better description of their data than the model given by Eq. (9). Note again that this response function has a non-zero value at $t = 0$. The Fourier transform of this response is given by:

$$\Gamma_4(\Delta\omega) = \frac{\tau_1}{1 + i\Delta\omega\tau_1} + \frac{C}{2i} \left[\frac{\tau_2}{1 + i(\Delta\omega - \frac{1}{\beta})\tau_2} - \frac{\tau_2}{1 + i(\Delta\omega + \frac{1}{\beta})\tau_2} \right]. \quad (14)$$

Displayed in Figure 5(a) is the fit of the square magnitude of this function to our spectral data. The average values of the parameters obtained from fits to our four scans are: $C = 7.45 \pm 3.96$, $\tau_1 = 1.15 \pm .078$ ps, $\tau_2 = 0.150 \pm 0.010$ ps, and $\beta = 0.274 \pm 0.096$ ps. The impulse response function given by Eq. (13), using these parameter values and normalized for unit height, is plotted in Figure 5(b) along with the results presented by

Nelson and co-workers.¹⁸ An examination of this figure and Table 3 indicates that, again, our results agree reasonably well with those reported by Nelson *et al.* The χ^2 values for the fits to this model range from 4.66 to 6.92. Although these values represent an average *decrease* of 11.7% relative to the χ^2 values obtained in fitting the overdamped model of Eqs. (9) and (10), they give an average *increase* of approximately 21.5% over the χ^2 values from the fits to the overdamped model of Eqs. (11) and (12). Thus we are led to the conclusion that while an underdamped model for the fast component gives a better fit than a similar overdamped model, it appears even more important to consider models which give a purely inertial response, i.e. that have a zero value at $t = 0$.

With this in mind, let us consider the following model which is zero at $t = 0$ and which also provides an underdamped response for the fast component:

$$\Delta n_5(t) = \frac{A}{2} \exp(-t/\tau_1)(1 - \exp(-2t/\beta)) + (1-A) \exp(-t/\tau_2) \sin(t/\beta), \quad t \geq 0, \quad (15)$$

and its Fourier transform which is given by:

$$\begin{aligned} \Gamma_5(\Delta\omega) = & \frac{A}{2} \left[\frac{\tau_1}{1 + i\Delta\omega\tau_1} - \frac{\gamma_1}{1 + i\Delta\omega\gamma_1} \right] \\ & + \frac{(1-A)}{2i} \left[\frac{\tau_2}{1 + i(\Delta\omega - \frac{1}{\beta})\tau_2} - \frac{\tau_2}{1 + i(\Delta\omega + \frac{1}{\beta})\tau_2} \right], \end{aligned} \quad (16)$$

where the parameter γ is defined to be:

$$\frac{1}{\gamma_1} = \frac{1}{\tau_1} + \frac{2}{\beta}. \quad (17)$$

Note that this model, despite its complexity, retains the same number of parameters as all of the other "inertial" models that have thus far been presented. The fit of this model to our data is shown in Figure 6(a), and a normalized plot of the impulse response described by Eq. (15), using the average parameters obtained from our curve-fits, is displayed in Figure

6(b). The average values that we obtain for these parameters are: $A = 0.221 \pm 0.084$, $\tau_1 = 1.09 \pm .07$ ps, $\tau_2 = 0.131 \pm 0.011$ ps, and $\beta = 0.230 \pm 0.098$ ps. If we define for Eq. (15) an "effective" $C = (1-A)/A$, an examination of Table 3 and a close inspection of Figures 5(b) and 6(b) indicate that these results are in very good agreement with those reported by Nelson *et al.*¹⁸ It is an interesting observation that, although there is no compelling physical basis for requiring the β 's which appear in the exponential and sine functions of Eq. (15) to be identical, a version of this model which provided two distinct β parameters produced absolutely no improvement in the fit. The χ^2 values of our fits to the model given by Eqs. (15) and (16) range from 3.09 to 5.12; on average this is a 36.2% decrease over the χ^2 values of the fits to the underdamped model of Eqs. (13) and (14), and a 22.4% decrease relative to the fits to the overdamped model given by Eqs. (11) and (12)—again a clear demonstration of the purely inertial character of the nuclear contribution to the optical Kerr response of CS₂.

As a final model, let us consider a three component orientational response that has been reported by Kenney-Wallace and co-workers.^{16,30} This model can be easily constructed from Eq. (15) with two additional parameters:

$$\begin{aligned} \Delta n_6(t) = & \frac{1}{2} (A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)) (1 - \exp(-2t/\beta)) \\ & + (1 - A_1 - A_2) \exp(-t/\tau_3) \sin(t/\beta), \quad t \geq 0. \end{aligned} \quad (18)$$

The Fourier transform of this expression is given by:

$$\begin{aligned} \Gamma_6(\Delta\omega) = & \frac{A_1}{2} \left[\frac{\tau_1}{1 + i\Delta\omega\tau_1} - \frac{\gamma_1}{1 + i\Delta\omega\gamma_1} \right] + \frac{A_2}{2} \left[\frac{\tau_2}{1 + i\Delta\omega\tau_2} - \frac{\gamma_2}{1 + i\Delta\omega\gamma_2} \right] \\ & + \frac{(1 - A_1 - A_2)}{2i} \left[\frac{\tau_3}{1 + i(\Delta\omega - \frac{1}{\beta})\tau_3} - \frac{\tau_3}{1 + i(\Delta\omega + \frac{1}{\beta})\tau_3} \right], \end{aligned} \quad (19)$$

where the parameters γ_i are given by:

$$\frac{1}{\gamma_1} = \frac{1}{\tau_1} + \frac{2}{\beta}, \text{ and } \frac{1}{\gamma_2} = \frac{1}{\tau_2} + \frac{2}{\beta}. \quad (20)$$

The results of our fit to this model are displayed in Figure 7(a), and the normalized impulse response function described by Eq. (18) is plotted in Figure 7(b) using the following average parameter values: $A_1 = 0.195 \pm 0.050$, $\tau_1 = 1.33 \pm .11$ ps, $A_2 = 0.278 \pm .075$, $\tau_2 = 0.420 \pm .117$ ps, $\tau_3 = 0.152 \pm 0.020$ ps, and $\beta = 0.154 \pm 0.031$ ps. Figure 7(b) and Table 4 indicate very good agreement between the impulse response parameter values we observe and those reported by Kenney-Wallace *et al.*^{17,30,35} The χ^2 values for the curve-fits of this model to our data range from 2.45 to 4.81. Again, because we have increased the number of parameters in the model, we must subject this reduction in χ^2 values to an F-test to determine the significance of the improvement in the goodness-of-fit. For two of the four scans, the F-test indicated acceptance of the more complicated model of Eq. (21) at a confidence level of 95% or better. The third scan produced an F-statistic indicating acceptance at the 90% confidence level, while the fourth scan showed only a small reduction in its χ^2 value and thus produced an F-statistic confidence level of about 70%, indicating a probability of approximately 30% that the χ^2 reduction occurred by random chance.

Summary and Conclusions

To summarize, we have measured the ultrafast response of the nuclear orientation component of the optical Kerr effect in CS₂ using a frequency domain technique and have fit our data to several simple models of this response. Although visually the differences between these response curves are quite subtle, we can nevertheless use the relative changes in χ^2 values³⁶ and, where appropriate, the F-test to observe general features indicated by the measured response curves and to effectively determine which model best

describes our data. With three of the four scans fit indicating an F-test confidence level of 90%, the three-component, underdamped model given by Eqs. (18) and (19) best describes our measured response curves. More importantly, the fits to the set of models presented here indicate the inertial nature, i.e. $\Delta n(t = 0) = 0$, of the nuclear component of the optical Kerr response of CS₂. From a fundamental physical point of view, this should not be surprising. Even in the absence of collisions which hinder the rotation of the molecules one would expect that the inertia of the molecules would prevent the creation of instantaneous alignment; only the electrons of the molecule are light enough to respond in near-instantaneous fashion.

Although the long-timescale component ($\tau \sim 1$ ps) of the nuclear response has been widely discussed in terms of the process of orientational diffusion,^{3,5-9,11-18,28,29,32-34} the nature of the subpicosecond response is a subject of much debate. Most often the fast dynamics have been discussed in terms of interaction-induced effects between partially aligned dipole moments,^{12,13,34} in terms of oscillatory or librational motion within intermolecular potentials in the liquid,^{4,17,18,29} and quite recently in terms of both of these phenomena.¹⁶ We do not advocate any particular physical interpretation for the specific functional forms given in Eqs. (5) – (19). In fact, the improvement in the degree to which these fits pass our goodness-of-fit tests does not increase nearly as dramatically as the reductions in χ^2 values might suggest. Despite the low χ^2 values, the fits of even the three-component model of Eqs. (18) and (19) fail the Durbin-Watson test³⁸ at the 1% confidence level,³⁹ and, on average, barely pass the runs up/down test²² at the 10% confidence level—in contrast, the fits to the measured responses of triphenyl methane dyes presented by Trebino, Barker, and Siegman¹⁹ easily passed these goodness-of-fit tests. The goodness-of-fit tests thus continue to suggest that these models are not wholly accurate descriptions of the responses that we are measuring in CS₂. Consequently, in the absence of a fundamental theory of the dynamics of the optical Kerr effect in liquids, we view these models as reasonable, empirical approximations which fit our data with varying degrees of

success. As such they provide estimates of the timescales of the nuclear motion of CS₂ molecules in the liquid state and indicate general features of the response—most notably the purely inertial nature of the nuclear orientation contribution to the ultrafast response of the optical Kerr effect.

Acknowledgements

We wish to thank Dr. D. J. McGraw for his assistance, Prof. G. A. Kenney-Wallace and Prof. K. A. Nelson for interesting discussions, and the Air Force Office of Scientific Research for support.

References

1. P. D. Maker, R. W. Terhune, and C. M. Savage, "Intensity-Dependent Changes in the Refractive Index of Liquids," *Phys. Rev. Lett.*, vol. 12, no. 18, pp. 507-509, May 4, 1964.
2. G. Mayer and F. Gires, "Action d'une onde lumineuse intense sur l'indice de refraction des liquides," *C. R. Acad. Sc. Paris*, vol. 258, pp. 2039-2042, 1964.
3. S. L. Shapiro and H. P. Broida, "Light Scattering from Fluctuations in Orientations of CS₂ in Liquids," *Phys. Rev.*, vol. 154, no. 1, pp. 129-138, 1967.
4. R. Cubeddu, R. Polloni, C. A. Sacchi, and O. Svelto, "Self-Phase Modulation and Rocking of Molecules in Trapped Filaments of Light with Picosecond Pulses," *Phys. Rev. A*, vol. 2, no. 5, pp. 1955-1963, 1970.
5. Y. Higashigaki, S. L. Whittenburg, and C. H. Wang, "Light scattering studies of orientational fluctuations of CS₂," *J. Chem. Phys.*, vol. 69, no. 7, pp. 3297-3301, 1978.
6. P. P. Ho and R. R. Alfano, "Optical Kerr effect in liquids," *Phys. Rev. A*, vol. 20, no. 3, pp. 2170-2187, 1979.
7. J-M. Halbout and C. L. Tang, "Femtosecond interferometry for nonlinear optics," *Appl. Phys. Lett.*, vol. 40, no. 9, pp. 765-767, 1982.
8. B. I. Greene and R. C. Farrow, "Femtosecond Transient Birefringence in CS₂," in *Picosecond Phenomena III*, eds. K. B. Eisenthal, R. M. Hochstrasser, W. Kaiser, and A. Lauberau, (Springer-Verlag, Berlin, 1982), pp. 209-211.
9. G. A. Kenney-Wallace and S. C. Wallace, "Picosecond Molecular Dynamics in Liquids via Degenerate Four-Wave Mixing," *IEEE J. Quantum Electron.*, vol. QE-19, pp. 719-723, Apr. 1983.
10. J. Etchepare, G. Grillon, I. Thomazeau, A. Migus, and A. Antonetti, "Third-order electronic susceptibilities of liquids measured by femtosecond kinetics of optical Kerr effect," *J. Opt. Soc. Am. B*, vol. 2, no. 4, pp. 649-653, 1985.
11. M. A. Duguay and J. W. Hansen, "An Ultrafast Light Gate," *Appl. Phys. Lett.*, vol. 15, no. 6, pp. 192-194, 1969.
12. G. A. Kenney-Wallace, "Non-Linear Optical Spectroscopy and Molecular Dynamics in Liquids," in *Applications of Picosecond Spectroscopy to Chemistry*, ed. K. B. Eisenthal, (D. Reidel, 1984), pp. 139-162.

13. P. A. Madden, "Interaction-Induced, Subpicosecond Phenomena in Liquids," in *Ultrafast Phenomena IV*, eds. D. H. Auston and K. B. Eisenthal, (Springer-Verlag, New York, 1984), pp. 244-251.
14. S. Kielich, "General Molecular Theory and Electric Field Effects in Isotropic Dielectrics," in *Dielectric and Related Molecular Processes*, Vol. 1, ed. M. Davies, (The Chemical Society, London, 1972).
15. C. Kalpouzos, W. T. Lotshaw, D. McMorrow, and G. A. Kenney-Wallace, "Femtosecond Laser-Induced Kerr Responses in Liquid CS₂," *J. Phys. Chem.*, vol. 91, pp. 2028-2030, 1987.
16. D. McMorrow, W. T. Lotshaw, and G. A. Kenney-Wallace, "Femtosecond Optical Kerr Studies on the Origin of the Nonlinear Responses in Simple Liquids," *IEEE J. Quantum Electron.*, vol. QE-24, pp. 443-454, Feb. 1988.
17. S. Ruhman, L. R. Williams, A. G. Joly, B. Kohler, and K. A. Nelson, "Nonrelaxational Inertial Motion in CS₂ Liquid Observed by Femtosecond Time-Resolved Impulsive Stimulated Scattering," *J. Phys. Chem.*, vol. 91, pp. 2237-2240, 1987.
18. S. Ruhman, B. Kohler, A. G. Joly, and K. A. Nelson, "Molecular Dynamics in Liquids from Femtosecond Time-Resolved Impulsive Stimulated Scattering," *IEEE J. Quantum Electron.*, vol. QE-24, pp. 470-481, Feb. 1988.
19. R. Trebino, C. E. Barker, and A. E. Siegman, "Tunable-Laser-Induced Gratings for the Measurement of Ultrafast Phenomena," *IEEE J. Quantum Electron.*, vol. QE-22, pp. 1413-1430, Aug. 1986.
20. A. E. Siegman, "Proposed picosecond excited-state measurement method using a tunable-laser-induced grating," *Appl. Phys. Lett.*, vol. 30, no. 1, pp. 21-23, 1977.
21. R. Trebino and A. E. Siegman, "Subpicosecond relaxation study of malachite green using a three-laser frequency-domain technique," *J. Chem. Phys.*, vol. 79, no. 8, pp. 3621-3626, 1983.
22. J. N. Demas, *Excited State Lifetime Measurements*, (Academic, New York, 1983).
23. J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308-313, 1965.
24. P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, (McGraw-Hill, New York, 1969).
25. A. G. Jacobson and Y. R. Shen, "Coherent Brillouin Spectroscopy," *Appl. Phys. Lett.*, vol. 34, no. 7, pp. 464-467, 1979.

26. C. E. Barker, "Ultrafast Nonlinear Optical Measurements Using the Tunable-Laser-Induced Grating Technique," Ph.D. dissertation, Stanford University, Stanford, CA, November 1988.
27. R. W. Hellwarth, "Third Order Optical Susceptibilities of Liquids and Solids," Prog. Quant. Electr., vol. 5, pp. 1 – 68, 1977.
28. K. Sala and M. C. Richardson, "Optical Kerr effect induced by ultrashort laser pulses," Phys. Rev. A, vol. 12, no. 3, pp. 1036-1047, 1975.
29. V. S. Starunov, "Scattering of Light Due to Anisotropy Fluctuations in Low-Viscosity Liquids," Opt. Spect. (USSR), vol. 18, pp.165-170, 1965.
30. C. Kalpouzos, W. T. Lotshaw, D. McMorro, and G. A. Kenney-Wallace, "Femtosecond Laser-Induced Optical Kerr Dynamics in CS₂/Alkane Binary Solutions," Chem. Phys. Lett., vol. 150, no. 1, 2, pp. 138–146, Sept. 1988.
31. P. Debye, *Polar Molecules*, (Chemical Catalog Co., New York, 1929).
32. B. J. Berne and R. Pecora, *Dynamic Light Scattering*, (Wiley, New York, 1976).
33. J. P. McTague, P. A. Fleury, and D. B. DuPre, "Intermolecular Light Scattering in Liquids," Phys. Rev., vol. 188, no. 1, pp 303-308, 1969.
34. D. Kivelson and P. A. Madden, "Light Scattering Studies of Molecular Liquids," Ann. Rev. Phys. Chem., vol.31, pp. 523–558, 1980.
35. C. Kalpouzos, D. McMorro, W. T. Lotshaw, and G. A. Kenney-Wallace, "Comment on: Femtosecond Laser-Induced Optical Kerr Dynamics in CS₂/Alkane Binary Solutions," to appear in Chem. Phys. Lett.
36. The χ^2 values reported here are consistently too large due to our inability to accurately account for all the noise sources in our measurements. While not affecting the utility of comparing relative values of χ^2 for different models, this situation does prevent the use of the χ^2 test as a measure of goodness-of-fit for any individual model.
37. D. McMorro, private communication.
38. N. Draper and H. Smith, *Applied Regression Analysis*, Second Edition, (Wiley, New York, 1981).
39. The confidence level indicated by these tests is the probability that the particular statistical characteristic calculated will occur by random chance in a normally-distributed set of random numbers possessing the same mean and standard deviation as the residuals from the fit. As such, these tests do not guarantee that a fit is "good," rather they provide evidence, or lack thereof, for rejection of a fit.

Tables

Table 1. Two-exponential decay model – Eq. (5)

	A	τ_1 (ps)	τ_2 (ps)
this work (TLIG)	0.36	1.21	0.18
Halbout and Tang ⁷	0.14	2.07	0.36
Greene and Farrow ⁸	0.35	2.16	0.24

Table 2. Overdamped inertial models – Eqs. (9), (11)

	B	τ_1 (ps)	τ_2 (ps)	τ_3 (ps)
this work (TLIG) – Eq. (9)	58.9	1.24	0.134	0.123
Nelson and co-workers ¹⁷ – Eq. (9)	44.0	1.37	0.114	0.102
this work (TLIG) – Eq. (11)	110 ^a	1.19	0.112	0.106

^a For the purposes of comparison an effective B is defined for Eq. (11) as $B = (1-A)/A$

Table 3. Underdamped inertial models – Eqs. (13), (15)

	C	τ_1 (ps)	τ_2 (ps)	β (ps)
this work (TLIG) – Eq. (13)	7.45	1.15	0.150	0.274
Nelson and co-workers ¹⁸ – Eq. (13)	3.57	1.01	0.124	0.159
this work (TLIG) – Eq. (15)	3.52 ^a	1.09	0.131	0.230

^a For the purposes of comparison an effective C is defined for Eq. (15) as $C = (1-A)/A$

Table 4. Underdamped inertial model (3 components) – Eq. (18)

	A_1	τ_1 (ps)	A_2	τ_2 (ps)	τ_3 (ps)	β (ps)
this work (TLIG)	0.195	1.33	0.278	0.420	0.152	0.154
Kenney-Wallace <i>et al</i> ¹⁶	0.118 ^a	1.61 ^a	0.121 ^a	0.50 ^a	0.140	0.150

^a These parameter values are appropriate to the functional form given by Eq. (9) of reference 30.^{35,37}

Figure Captions

Figure 1 Simplified schematic of the Tunable-Laser-Induced Grating apparatus. The system consists essentially of three pulsed dye lasers pumped by a frequency-doubled, Q-switched YAG laser.

Figure 2 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{yxyx}$ of CS₂ fit to a two exponential decay model given by Eqs. (5) and (6). The parameter values shown represent the average values of fits to four such scans. (b) The two exponential decay impulse response given by Eq. (5) is shown for the best fit parameters obtained from our experiments and for the results reported by Halbout and Tang⁷ and Greene and Farrow.⁸

Figure 3 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{yxyx}$ of CS₂ fit to the overdamped model given by Eqs. (9) and (10). The parameter values shown represent the average values of fits to four such scans. (b) The impulse response given by Eq. (9) is shown for the best fit parameters obtained from our experiments and for the results reported by Nelson *et al.*¹⁷ Note that the impulse response is nonzero at $t = 0$.

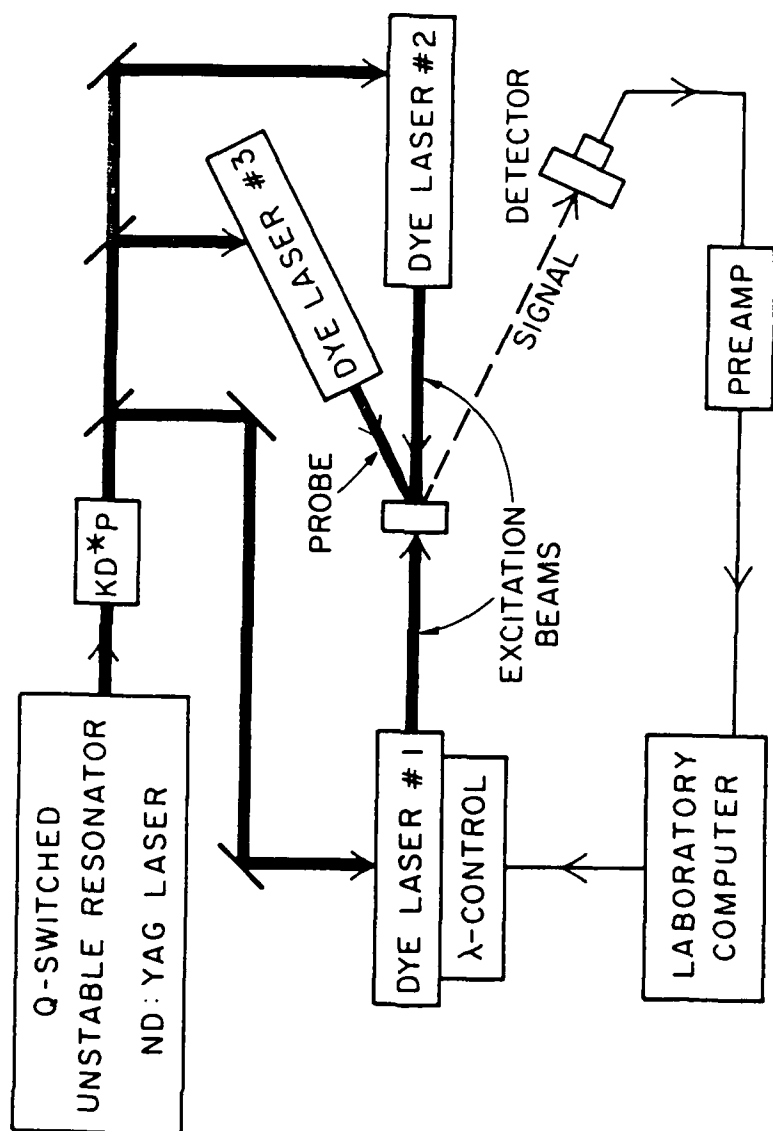
Figure 4 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{yxyx}$ of CS₂ fit to the overdamped model given by Eqs. (11) and (12). The parameter values shown represent the average values of fits to four such scans. (b) The overdamped impulse response given by Eq. (11) is shown for the best fit parameters obtained from our experiments. Note that this response is zero at $t = 0$.

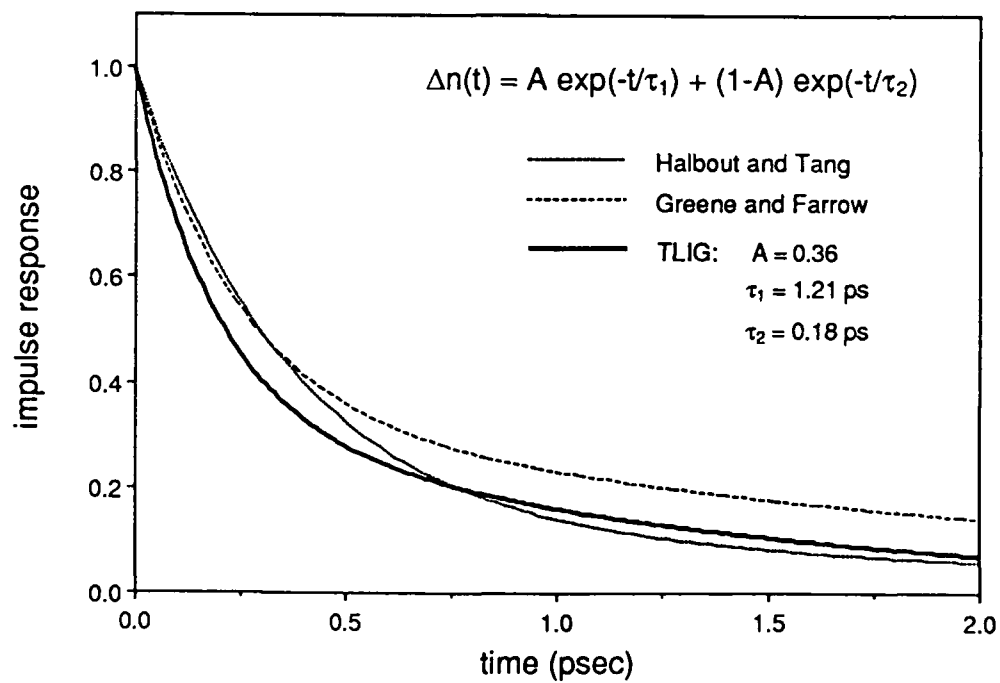
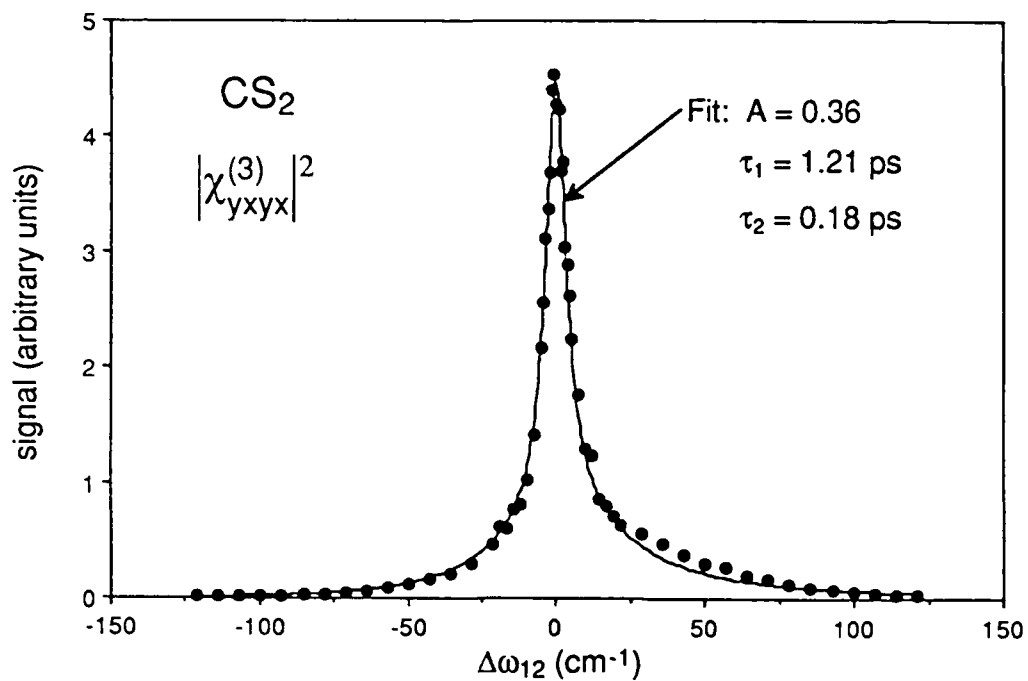
Figure 5 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{yxyx}$ of CS₂ fit to the underdamped model given by Eqs. (13) and (14). The parameter values shown represent the average values of fits to four such scans. (b) The underdamped impulse response given by Eq. (13) is shown for the best fit parameters obtained from our experiments and for the results reported by Nelson *et al.*¹⁸ Note that the impulse response is nonzero at $t = 0$.

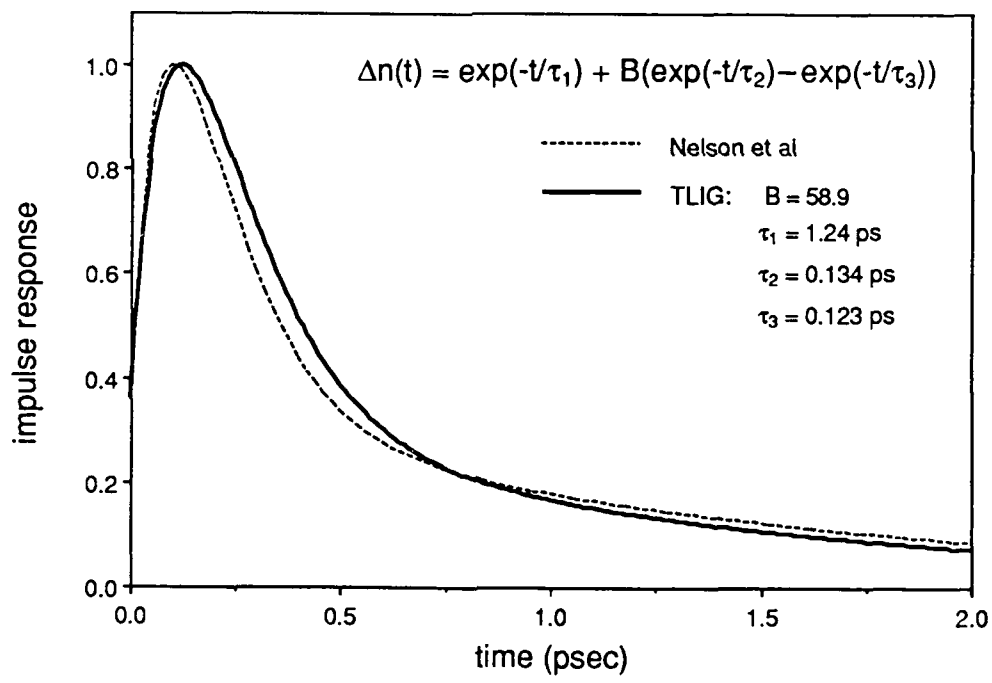
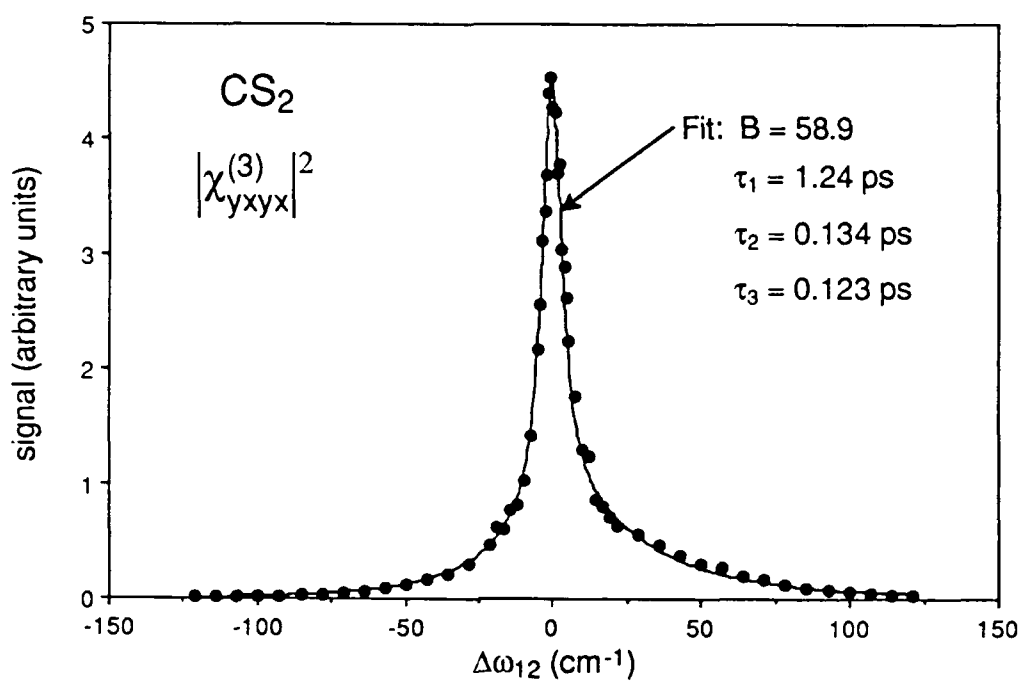
Figure 6 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{yxyx}$ of CS₂ fit to the underdamped model given by Eqs. (15) and (16). The parameter values shown represent the average values of fits to four such scans. (b) The underdamped impulse response given by Eq. (15)

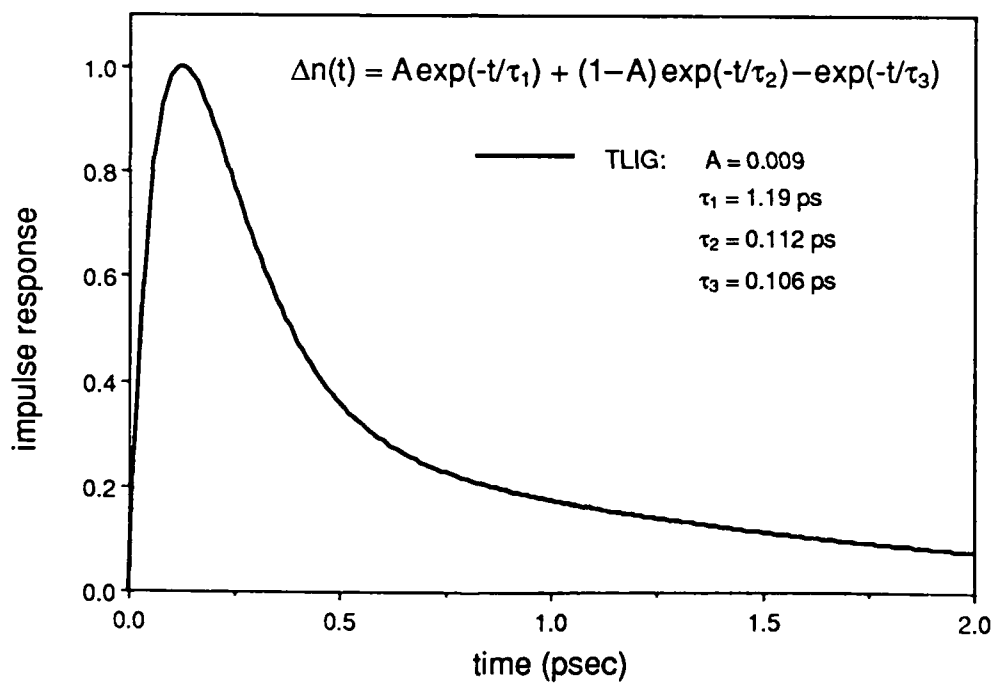
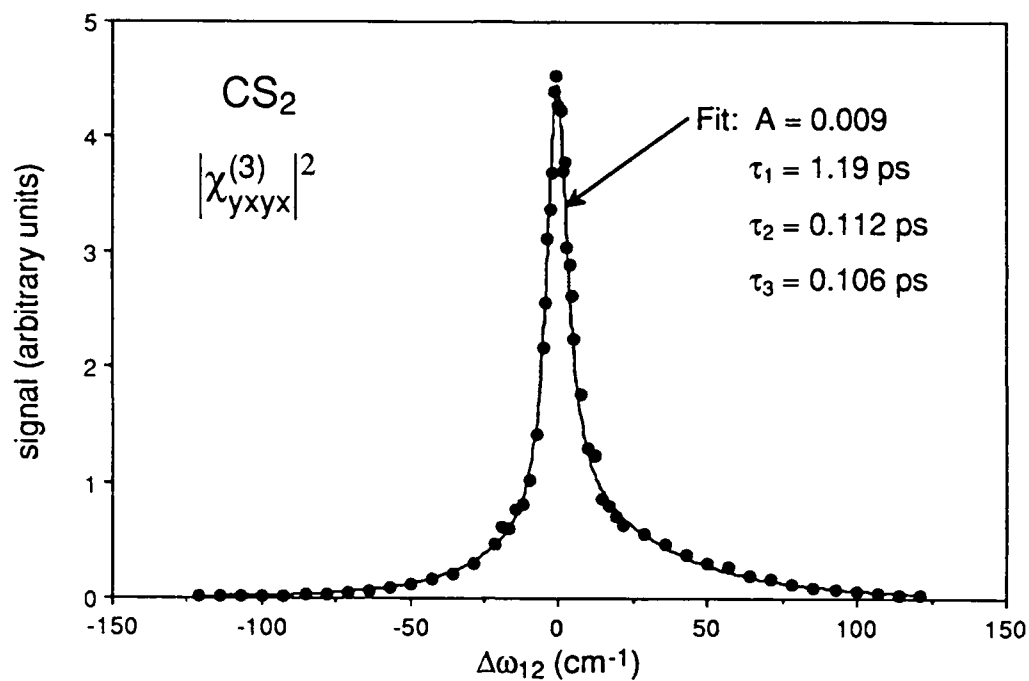
is shown for the best fit parameters obtained from our experiments. Note that the impulse response is zero at $t = 0$.

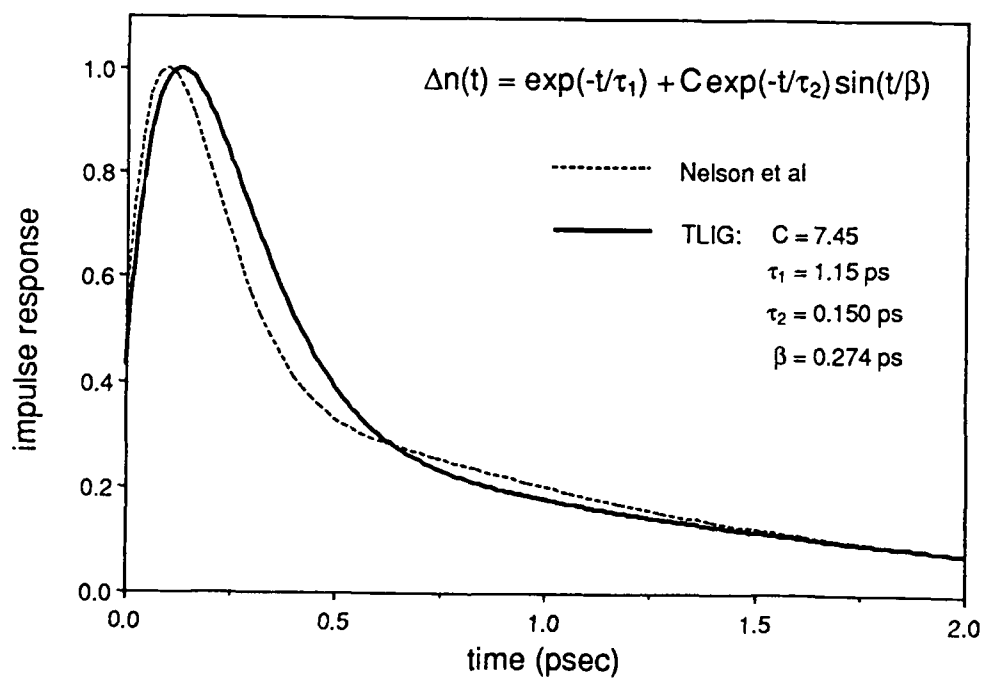
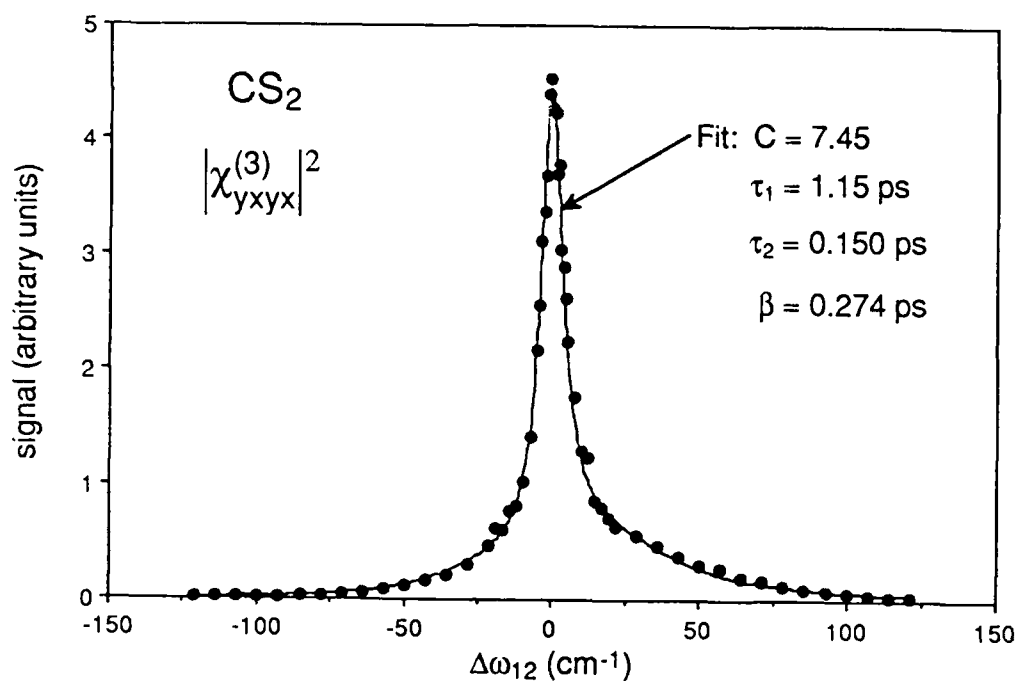
Figure 7 (a) The nonlinear susceptibility tensor $\chi^{(3)}_{xyx}$ of CS_2 fit to the underdamped model given by Eqs. (18) and (19). The parameter values shown represent the average values of fits to four such scans. (b) The underdamped impulse response given by Eq. (18) is shown for the best fit parameters obtained from our experiments and for the results reported by Kenney-Wallace *et al.*^{16,35,37} Note that the impulse response is zero at $t = 0$.

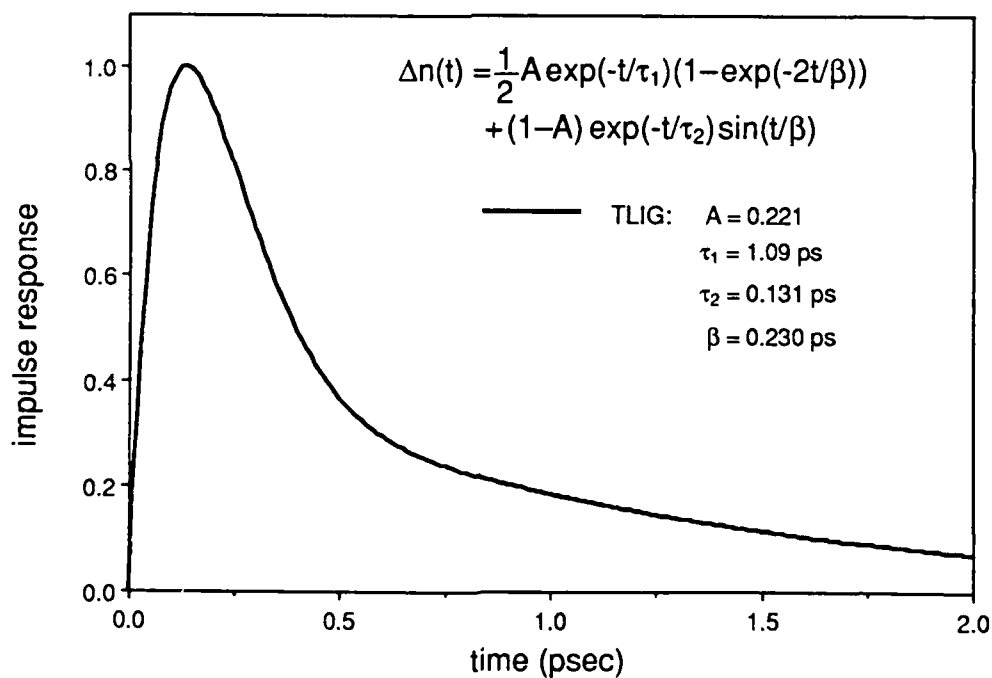
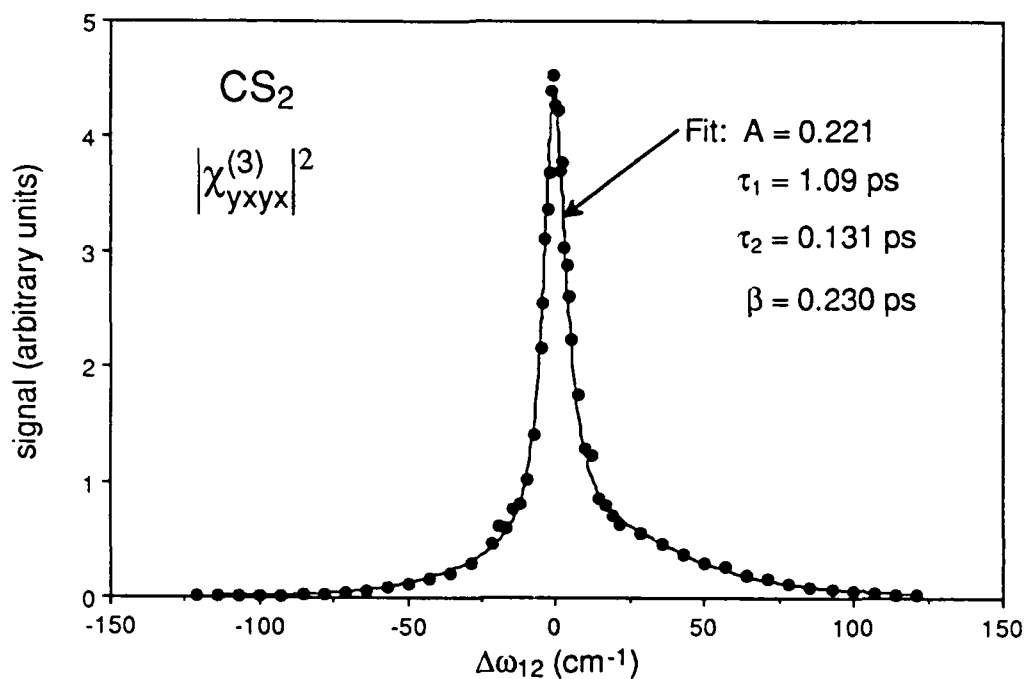


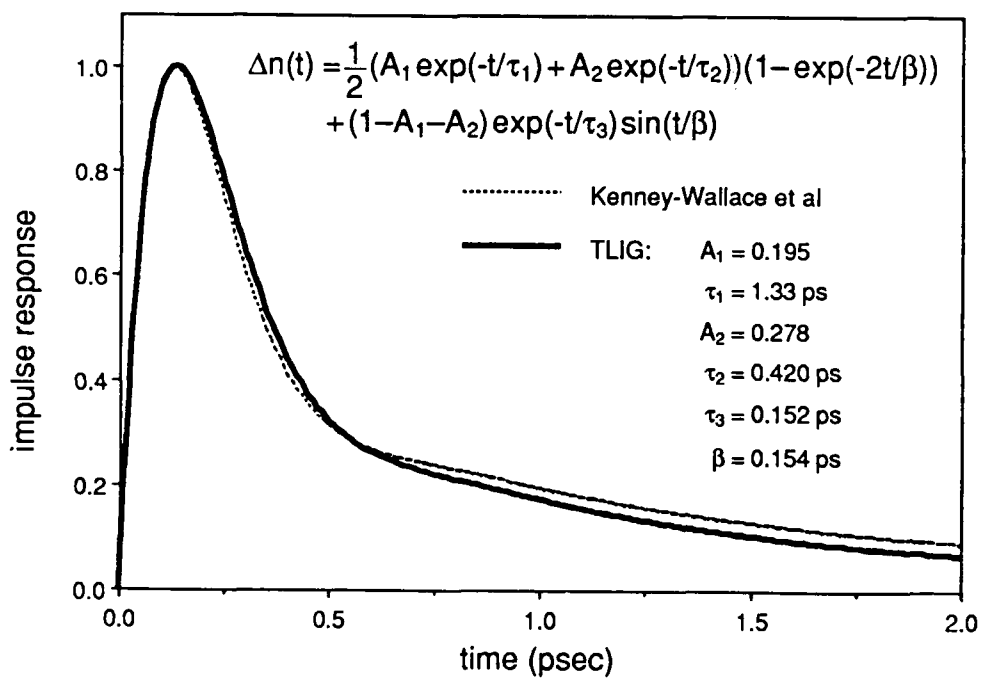
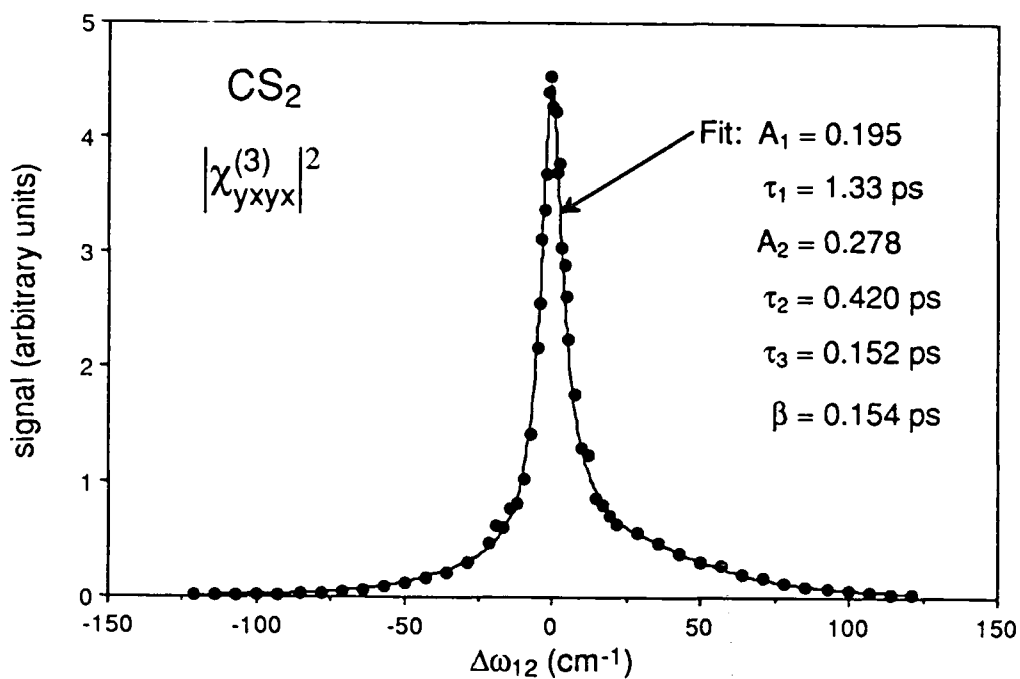












Resolution of the nuclear and electronic contributions to the optical nonlinearity in polysilanes

Daniel J. McGraw^{a)} and A. E. Siegman

E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305

G. M. Wallraff and R. D. Miller

IBM Almaden Research Center, 260 Harry Road, San Jose, California 95193

(Received 8 July 1988; accepted for publication 13 February 1989)

We present a method of resolving the nuclear and electronic contributions to the third-order susceptibility χ^3 in a phase conjugation experiment. The technique is applied to octylmethylpolysilane yielding χ^3 (electronic) = $(1.8 \pm 0.5) \times 10^{-12}$ esu and χ^3 (nuclear) = $(1.1 \pm 0.4) \times 10^{-12}$ esu at 532 nm.

Organic polymers possessing a conjugated backbone such as polyacetylene and polydiacetylene have attracted interest due to their potential for applications in nonlinear optics. Extended π electron delocalization creates a large anharmonic component in the nonresonant electron oscillations.¹ These organic semiconductors have been found by third-harmonic generation to have third-order optical susceptibilities (χ^3) as large as 10^{-8} esu.²⁻⁴

Polysilane, a silicon backbone polymer, has recently been recognized as a member of this class of backbone conjugated polymers. Although it has a sigma-bonded framework, there is sufficient orbital overlap to be referred to as a sigma-delocalized system.⁵ A number of soluble materials have been prepared and the polysilanes have been studied extensively as a photoresist for UV lithography.⁶⁻¹¹ Kazjar *et al.* have recently measured χ^3 in phenylmethylpolysilane in a third-harmonic generation experiment, finding χ^3 (1.06 nm) = 1.5×10^{-12} esu.¹² Baumert *et al.* have used third-harmonic generation to probe sidechain crystallization which results in backbone conformational changes in semicrystalline polydi-*n*-hexylsilane.¹³ The recent Kerr gate experiments of Yang *et al.* put an upper limit of three picoseconds on the time response of χ^3 in polysilane.¹⁴ The degenerate four-wave mixing experiments presented here measure contributions to χ^3 on all time scales, and separate the electronic and nuclear contributions to χ^3 . The separation of nuclear and electronic contributions is based on their different spatial symmetries not on assumptions of the decay time of the nuclear response.

The linear and nonlinear optical properties of polysilane are dominated by the silicon backbone. The polysilane derivatives are transparent throughout the visible, but absorb strongly from 300 to 400 nm depending on the substituents. Aromatic substituent groups attached to each silicon are weakly coupled to the backbone, causing a red shift in the lowest electronic $\sigma-\sigma^*$ transition of the backbone.^{15,16} The polysilane employed here, poly(*n*-octylmethylsilane; OMPS), is an amorphous fluid with the viscosity of rubber cement. The large anisotropic polarizability of the polysilane and its isotropic fluid structure allow the possibility of an orientational Kerr effect in polysilane.

The phase conjugation experimental configuration is

shown in Fig. 1. A Q-switched and frequency-doubled Nd:YAG laser provides 10 ns, 532 nm pulses at a 10 Hz repetition rate. The intensity was varied without changing the pulse shape by rotating the first of two glan polarizing prisms. The maximum pulse energy was 20 mJ in the counterpropagating pump beams and 5 mJ in the probe beam. The spot size of the pump beam is about 1.5 mm which yields a peak intensity of 30 MW/cm² at the sample. The sample of the neat OMPS was contained in 1 mm quartz cuvette. The phase conjugate signal (counterpropagating with respect to the probe) and the incident intensity are detected by silicon photodiodes and displayed on an oscilloscope.

Figure 2 is a plot of phase conjugate signal intensity versus pump intensity. The cubic power dependence confirms a third-order optical nonlinearity which can be due to either electronic, orientational, acoustic, or thermal processes. The phase conjugate beam faithfully reproduced the spatial profile of the doughnut mode probe beam. The maximum phase conjugation reflectivity observed was 0.2%. The phase conjugate intensity was calibrated by replacing the OMPS with carbon disulphide in an identical cell. The OMPS was found to have a χ^3 of $(2.9 \pm 0.5) \times 10^{-12}$ esu, 1.7 times greater than CS₂.¹⁷

The range of pump beam intensities over which a phase conjugate signal could be detected was about one order of magnitude. This was limited on the low end by Rayleigh scattering and on the high end by a single-shot damage

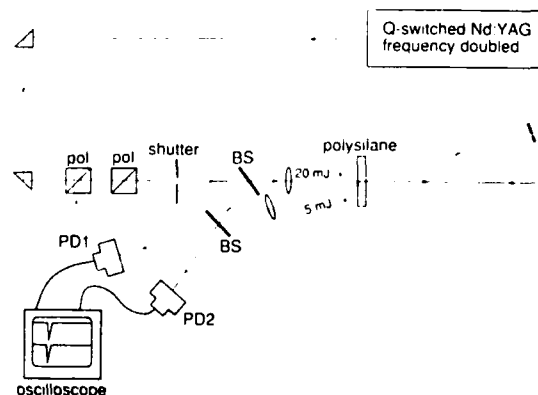


FIG. 1. Phase conjugation experimental schematic. BS, beamsplitter; PD1, phase conjugation signal detector; PD2, incident intensity photodiode. A half-wave plate is inserted in the probe beam path for perpendicular polarization.

^{a)} Current address: Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131.

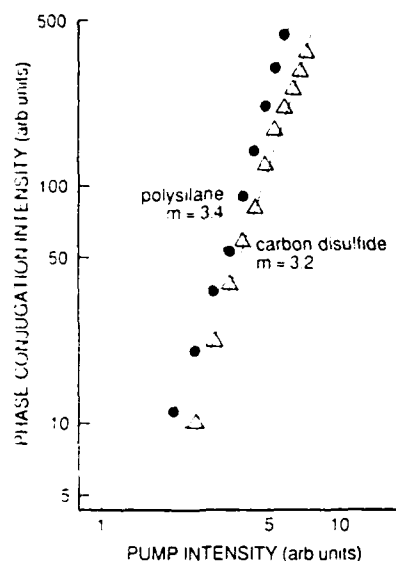


FIG. 2. Phase conjugation intensity vs pump intensity, displaying third-order power dependence and relative signal strength for identical 1 mm pathlength cells of poly(*n*-octylmethylsilane) and carbon disulphide. Carbon disulphide provides calibration of polysilane χ^3 measurement.

threshold of 100 MW/cm² or an incident energy flux of about 10 J/cm². The nature of the damage mechanism is not known. The polysilane one-photon absorption is too low ($< 10^{-4}$) to cause the damage observed. A two-photon absorption observed recently in polysilane is not seen here.¹⁸ Two-photon absorption-induced gratings would yield a fifth-order power dependence of the phase conjugation intensity on the incident intensity. The most likely cause of the damage is impurity absorption.

The nonlinearity shown in Fig. 2 for both OMPS and CS₂ is due to both electronic and orientational factors. Both molecules have an anisotropic linear polarizability, therefore they experience a torque tending to align the more polarizable axis of the molecule with the polarization plane of the light. The relative magnitude of the electronic hyperpolarizability and the orientational (or nuclear) contribution to χ^3 can be determined from dc Kerr effect and depolarized Rayleigh scattering experiments. This has been done for CS₂ which was found dominated by the orientational effect; $\chi^3(\text{orientational})/\chi^3(\text{orientational}) + \chi^3(\text{electronic}) = 0.89 \pm 0.01$.¹⁷ We report here (i) the first measurement of χ^3 for OMPS in a phase conjugation experiment, and (ii) the first separation of orientational and electronic contributions to χ^3 in a phase conjugation experiment by varying the probe polarization. This separation is possible due to the lower symmetry of the orientational nonlinearity compared with the electronic hyperpolarizability.

In isotropic media the four independent χ^3 tensor elements reduce to three by the isotropy relation: $\chi^3_{xxxx} = \chi^3_{xxyy} + \chi^3_{xyxy} + \chi^3_{yyxx}$, where by convention the order of the indices corresponds to diffracted field, pump No. 1, pump No. 2, and probe field.^{17,19} These three elements can be further reduced to one parameter if the nature of the nonlinearity is known. If the response is purely orientational the relative magnitude of the three elements becomes $\chi^3_{xxxx} = 1$, $\chi^3_{xyxy} = -1/2$, and $\chi^3_{xxyy} = \chi^3_{xyxy} = 3/4$.

These values result only from our assumptions of an

incompressible fluid in three dimensions. The χ^3 (orientational) ratios are the same for both rigid rod molecules like CS₂ and long-chain molecules like polysilane. These χ^3 ratios follow from the fact that the first-order change in the isotropic distribution of molecular orientations caused by alignment with a linearly polarized pump field is given by $\Delta\rho(\theta)\alpha Y_{20} = 3/2 \cos^2\theta - 1/2$, where θ is the polar angle. We see that this determines the ratio of the amplitude of the index gratings probed perpendicular and parallel to the pump fields: $\Delta n(\theta = \pi/2)/\Delta n(\theta = 0) = \chi^3_{xyxy}/\chi^3_{xxxx} = \Delta\rho(\theta = \pi/2)/\Delta\rho(\theta = 0) = -1/2$. This is a consequence of Poisson's law: the ratio of strains perpendicular and parallel to a uniaxial stress is $-1/2$ for a constant volume fluid. The other tensor elements follow from the isotropy relation and the fact that χ^3 (orientational) is symmetric about an interchange of the pump fields.

For purely electronic, instantaneous hyperpolarizability the above picture of a two-step process involving an induced material disturbance followed by diffraction from the induced grating is not correct. All three fields interact simultaneously in the medium so χ^3 must be symmetric about an interchange of any of these fields. This together with the isotropy relation yields $\chi^3_{xxxx} = 1$ and $\chi^3_{xxyy} = \chi^3_{xyxy} = \chi^3_{xyyx} = 1/3$. We would expect then that upon changing the probe from a parallel to a perpendicular polarization with respect to the pump beams the phase conjugation intensity should drop by a factor $(\chi^3_{xyxy}/\chi^3_{xyxy})^2 = 1/9$ for a purely electronic hyperpolarizability and 9/16 for a purely orientational nonlinearity.

Figure 3 shows the results of such an experiment. For OMPS the phase conjugation signal drops by a factor of 6.1 ± 1.2 which yields a ratio $\chi^3(\text{orientational})/\chi^3(\text{orientational}) + \chi^3(\text{electronic}) = 0.38 \pm 0.15$. Repeating this with CS₂ yields $\chi^3(0)/[\chi^3(0) + \chi^3(e)] = 0.86 \pm 0.05$, in good agreement with the accepted value of 0.89 ± 0.01 .¹⁷

We note here that any contribution from absorption-

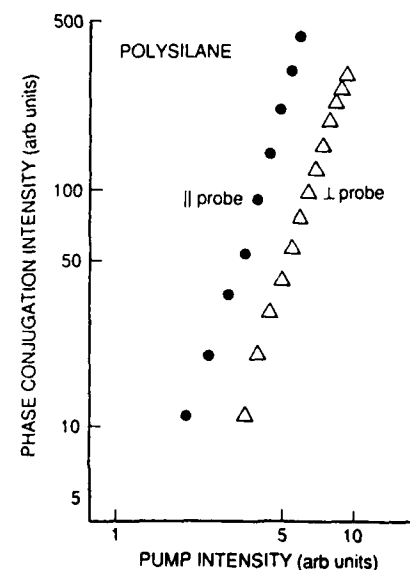


FIG. 3. Phase conjugation intensity vs pump intensity from poly(*n*-octylmethylsilane) with probe beam polarized parallel (dots) and polarized perpendicular (triangles) to the pump beams. The relative phase conjugation signal strengths yield the relative magnitude of the nuclear and electronic contributions to χ^3 .

induced thermal gratings will contribute to the phase conjugation signal when the probe beam is polarized parallel (PCPAR) but not when the probe is polarized perpendicular (PCPER) to the pump beams. If thermal gratings are present they will therefore introduce an error in our determination of the relative strengths of the electronic and orientational contributions to χ^3 in the direction of enhancing the apparent electronic contribution. While we cannot rule out the presence of a small thermal grating contribution, the observation that $\text{PCPER} > \text{PCPAR}/9$ can be accounted for only by a nuclear contribution to χ^3 . No combination of electronic χ^3 and thermal or any scalar gratings alone can account for the large phase conjugation signal observed with the probe polarized perpendicular to the pump beams.

That the nonlinear optical response is one-third nuclear and consequently slower than a purely electronic response limits optical switching speeds in device applications. As OMPS is a polymer and it assumes a coil conformation in the bulk the orientational motion it undergoes is not one of orientational diffusion of rigid rods as in CS_2 but is caused by a squeezing of the coiled silicon backbone about the polarization direction. The relaxation time is therefore expected to be approximately the silicon-silicon bending mode vibrational damping time. From Raman linewidths this is estimated to be on the order of a picosecond.²⁰

The three techniques applied to polysilane measurements: harmonic generation, time-resolved four-wave mixing, and the polarization discrimination four-wave mixing technique are complementary. The three methods report the same magnitude of the electronic contribution to χ^3 . The harmonic generation experiment is not sensitive to the nuclear response. The time-resolved χ^3 measurements could not distinguish between the nuclear and electronic contributions to the sub 3 ps response. The polarization discrimination technique presented here resolves the nuclear and electronic contributions to χ^3 based on their different spatial symmetries without assumptions on the time response of the nuclear contribution.

In conclusion, we have presented the first observation of phase conjugation in a polysilane and separated the nuclear and electronic contribution to χ^3 .

This research was supported by the Air Force Office of Scientific Research. One of us (D.J.M.) would like to thank Ad Kostenbauder for many illuminating discussions and Frank Schellenberg for inciting my interest in polysilanes.

- ⁹G. P. Agrawal, C. Cojan, and C. Flytzanis, *Phys. Rev. B* **17**, 776 (1978).
- ¹⁰W. Fan, S. Etemad, G. Baker, S. Benson, and J. Madey, presented at the APS Annual Meeting, New Orleans, LA, 1988.
- ¹¹G. M. Carter, J. V. Hryniewicz, M. R. Thakur, Y. J. Chen, and S. E. Meyler, *Appl. Phys. Lett.* **49**, 998 (1986).
- ¹²B. I. Green, J. Orenstein, R. R. Millard, and L. R. Williams, *Phys. Rev. Lett.* **58**, 2750 (1987).
- ¹³K. A. Klingensmith, J. W. Downing, R. D. Miller, and J. Michl, *J. Am. Chem. Soc.* **108**, 7438 (1986).
- ¹⁴R. West, L. D. David, P. I. Djurovich, K. L. Stearley, K. S. V. Srinivasan, and H. Yu, *J. Am. Chem. Soc.* **103**, 7352 (1981).
- ¹⁵For a recent review of polysilanes, see R. West, *J. Organomet. Chem.* **300**, 329 (1986).
- ¹⁶D. C. Hofer, R. D. Miller, C. G. Willson, and A. R. Neureuther, *Proc. SPIE* **469**, 108 (1984).
- ¹⁷D. C. Hofer, R. D. Miller, and C. G. Willson, *Proc. SPIE* **469**, 16 (1984).
- ¹⁸J. M. Zeigler, L. A. Harrah, and A. W. Johnson, *SPIE Advances in Resist Technology and Processing II*, 167 (1985).
- ¹⁹R. D. Miller, in *Silicon Chemistry*, edited by E. R. Corey, J. Y. Corey, and P. P. Gaspar (Ellis Horwood, Chichester, England, 1988), Chap. 35.
- ²⁰F. Kajzar, J. Messler, and C. Rossio, *J. Appl. Phys.* **60**, 3040 (1986).
- ²¹J. C. Baumert, D. H. Jundt, M. C. Jurich, H. Looser, R. D. Miller, J. Rabolt, S. Sooriyakumaran, J. D. Swalen, and R. J. Tveig, *Appl. Phys. Lett.* **53**, 1147 (1988).
- ²²L. Yang, Q. Z. Wang, P. P. Ho, R. Dorsinville, R. R. Alfano, W. K. Zou, and N. L. Yang, *Appl. Phys. Lett.* **53**, 1245 (1988).
- ²³J. Michl, J. W. Downing, T. Karatsu, K. A. Klingensmith, G. M. Wallraff, and R. D. Miller, *Organic and Organometallic Polymers*, ACS Symp. Series No. 360 (American Chemical Society, Washington, DC, 1988), Chap. 5.
- ²⁴L. A. Harrah and J. M. Zeigler, *Macromolecules* **20**, 601 (1987).
- ²⁵R. W. Hellwarth, *Prog. Quantum Electron.* **5**, 1 (1977).
- ²⁶F. M. Schellenberg, R. L. Byer, R. D. Miller, and R. Sooriyakumaran, presented at the International Conference on Quantum Electronics, Tokyo, paper TH14, 1988 (unpublished).
- ²⁷J. Etchepare, G. Grillion, I. Thomazeau, G. Hamoniaux, and A. Orszag, in *Ultrafast Phenomena V*, edited by G. Fleming and A. E. Siegman (Springer, Berlin, 1985), p. 504.
- ²⁸J. F. Rabolt, D. C. Hofer, R. D. Miller, and G. N. Fickes, *Macromolecules* **19**, 611 (1986).

High-speed diffusion-driven photodetector

Adnah G. Kostenbauder

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305

(Received 22 May 1987; accepted for publication 14 August 1987)

We describe a novel ultrafast photodetector based upon combining the Dember effect and optically generated fine carrier gratings. We present experimental data on prototype devices, and demonstrate that this device's response time is limited neither by the carrier lifetime nor by the saturation-velocity transit time.

Previous photodetectors have obtained high-speed operation either by using small geometries to reduce the electron transit time¹ or by relying upon shortened carrier lifetimes to rapidly terminate the photocurrent.²⁻⁴ In this letter we describe a new kind of high-speed photodetector which is based upon the rapid decay of photocarrier gratings and the associated Dember-effect potentials.

In its simplest form this detector consists of a slab of semiconducting material with identical transparent ohmic contacts on two opposing faces. Assume this slab is illuminated by the same light beam normally incident through both faces, so that the resulting optical interference pattern generates an electron-hole plasma in a grating pattern within the semiconductor. For maximum signal, one electrode must be located at a minimum of this pattern, while the other should be at a peak. Alternatively, the electron-hole grating may be obtained by imaging a periodic reticle onto an end face of the slab. The presence of this asymmetric hole-electron grating then produces a photovoltaic potential across the slab, between the two electrodes. The only symmetry-breaking element in this device is the carrier grating. As a result, the photogenerated potential can only last as long as the carrier grating. Within reasonable design constraints this photoresponse can have picosecond time constants and quantum efficiencies near unity.

To understand the operation of this photodetector, its open-circuit voltage, as a function of time after an optical impulse is received, can be calculated from the well known dynamic equations for carrier densities.⁵ Under the assumption that the injected densities are $N(t=0) = P(t=0) = B + C \cos(kx)$ with $B \gg C$, and that the dielectric relaxation time associated with each of the two carriers is fast compared to the diffusion times, we obtain

$$N = P = B \exp(-t/\tau) + C \cos(kx) \times \exp[-(D_a k^2 + 1/\tau)t], \quad (1)$$

where $D_a = 2D_n D_p / (D_n + D_p)$ is the ambipolar diffusion constant and τ is the effective carrier lifetime. The potential associated with this is then well approximated by

$$V = \frac{KT}{q} \frac{D_n - D_p}{D_n + D_p} \frac{C \cos kx}{B} \exp\left[-\left(D_a k^2 + \frac{1}{\tau}\right)t\right], \quad (2)$$

and the open-circuit potential between the electrodes is just this same expression with $\cos kx$ replaced by 2.

In physical terms, the electron and hole carrier gratings try to flatten out at different speeds due to their different

diffusion coefficients. After a few dielectric relaxation times, the space charge created by the separation of electrons and holes creates an electric field which acts to prevent further separation. The potential associated with this space charge (the Dember voltage⁶) tracks the decaying amplitude of the grating, and provides the photovoltaic output of the device.

The above analysis assumes no junction fields near the electrodes. Intuition suggests that such fields might destroy the simple grating decay picture, but this is not the case. The generation of a potential difference between the two ends requires a symmetry-breaking element, and the device is symmetrical except for the grating. Thus, the generated potential decays with the grating even in the presence of junction fields. Also, because a small fractional separation of the injected electrons and holes can cancel any small junction fields in the device, one expects that the time constant will only be weakly dependent on the field. It is not possible to obtain analytic solutions for the potential in the presence of junction fields, but computer simulations with silicon's material parameters indicate that the photovoltage's time constant is changed by less than a factor of 2 provided that the junction fields are less than about a thermal voltage per grating period.

We can also estimate the quantum efficiency of this device (into a short-circuit load) from Eq. (2) and the internal conductance of the detector. For a high modulation grating in the regime where the response time is dominated by diffusion, $\eta \approx |D_n/D_p - D_p/D_n|(kL)^{-2}$, which can be reasonably high in materials such as GaAs where D_n/D_p is large and kL is a few π .

To demonstrate this concept experimentally, a number of prototype detectors were fabricated from 100- μm -thick, high-resistivity ($N = 2 \times 10^{13} \text{ cm}^{-3}$) silicon wafers by heavily doping both surfaces and then depositing about 1 μm of aluminum. These wafers were immersed in LN_2 , cleaved into rectangles from 1 to 2 mm on a side, and the pieces sorted to find ones with a flat facet. Acceptable pieces were bonded to the back side of SMA coaxial-connector flanges with a fine wire connection from one face to the coaxial center conductor, and with a flat cleaved surface protruding beyond the flange to allow optical access.

These samples were excited by a synchronously pumped mode-locked dye laser operating at 590 nm. The pulses had a duration of 5 ps, energies of 10 nJ, and a repetition rate of 150 kHz. The pulse train was expanded and transmitted through a Ronchi ruling which was then imaged onto the cleaved edge facet of the detector. By using rulings with various per-

iods and microscope objectives with different powers, it was possible to produce fringe patterns with a wide variety of fringe spacings on the detector edges. The output of the detector was connected directly to an S-4 sampling head.

Since the photovoltaic signal depends on the position of the photocarrier grating relative to the electrodes, moving the fringe pattern transversely by one-half of its period should reverse the polarity of the signal, and this was in fact observed. This signature is characteristic only of grating induced currents and thus eliminates the possibility that we were measuring a photoconductive signal due to a stray bias voltage.

Typical signal levels ranged from 2 to 5 mV. Oscilloscope photographs of these signals were digitized and fit to single exponentials. Figure 1 shows the dependence of the time constant on the grating spacing ($\Lambda = 2\pi/k$) for three different detectors. The ambipolar diffusion coefficient determined from these measurements is $7.0 \pm 2.0 \text{ cm}^2/\text{s}$ and the effective carrier lifetime of $400 \pm 150 \text{ ns}$. Given the uncertainties introduced by junction fields, this diffusion constant is in good agreement with the silicon surface ambipolar diffusion constant, and the carrier lifetime is indicative of a surface recombination velocity of about 1000 cm/s . (The bulk carrier lifetime in this silicon was about $400 \mu\text{s}$ before fabrication of the contacts.) The error bars in Fig. 1 are primarily in grating spacing, rather than response time, due to the fact that a simple grating is imaged at a variety of planes behind the lens, but with a period which varies with position. Since the distance from the rear focal plane of the microscope objective to the detector is hard to set accurately, errors in fringe spacing are expected to be about 15%. Using this fact, we have produced response times as short as 350 ps

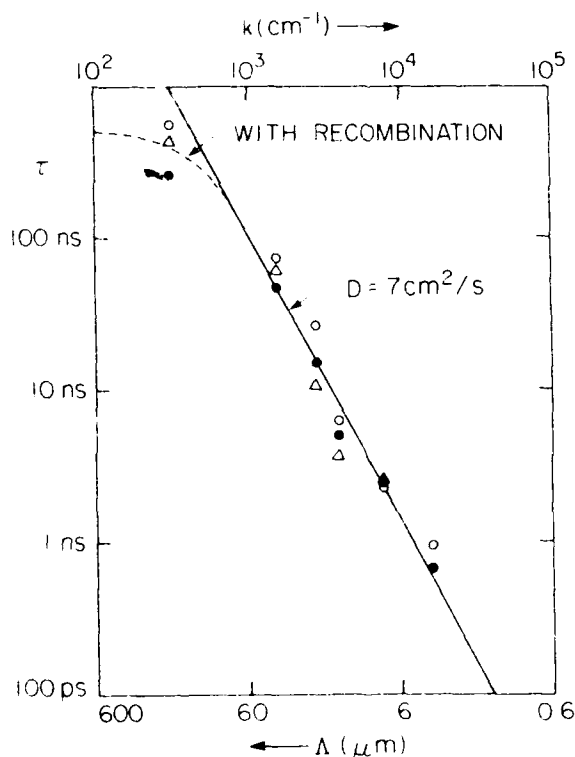


FIG. 1 Measured response times of three different $100\text{-}\mu\text{m}$ -thick silicon diffusion-driven detectors as a function of grating period.

by intentionally moving the microscope objective towards the detector. This time constant, along with the fastest ones shown in Fig. 1, is substantially below the saturation velocity transit time of $\approx 1.5 \text{ ns}$ for a $100\text{-}\mu\text{m}$ silicon device. Since the carrier lifetime at the surface is several hundred nanoseconds, we also have confirmation that this type of detector is not governed by any carrier lifetime limit.

Our theory for junction field-free devices predicts that the quantum efficiency scales as k^{-2} . This scaling law has been experimentally verified, although the proportionality constant is not well determined due to geometrical losses. Including these losses, the measured quantum efficiency is about 10^{-4} at $k = 4000 \text{ cm}^{-1}$. Even if the proportionality constant was determined, it would not be expected to be $[(D_n/D_p) - (D_p/D_n)]L^{-2}$ since computer simulations have shown that junction fields can significantly alter the magnitude of the current without altering the time constant.

Finally, we have measured the output voltage versus the input pulse energy for the $k = 1600 \text{ cm}^{-1}$ case, with results as shown in Fig. 2, indicating a linear response at low energies with the output saturating at 14 mV at large pulse energies. This is not in agreement with the expected 24-mV saturation voltage for high contrast gratings written in a junction-potential-free device. This discrepancy is believed to be caused by some combination of the device's junction fields, the modulation transfer function of the microscope objective, and the contact resistance.

In conclusion, we have demonstrated a novel type of high-speed photodetector whose response time is limited by neither its transit time nor its carrier lifetime. Experimental data have confirmed the expected dependence of the response time and the quantum efficiency on the grating fringe spacing. The predicted saturation has been observed, but at a value about half that predicted by our analytic theory. By constructing devices in which antiparallel optical beams are used to produce the carrier gratings, and by using GaAs in

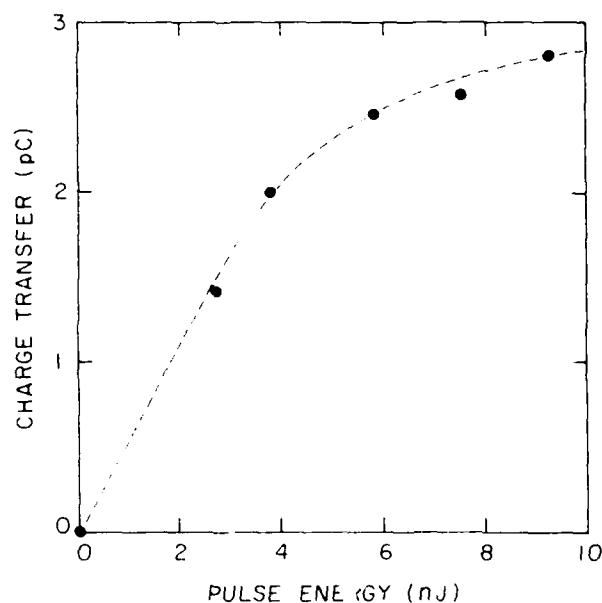


FIG. 2 Charge transfer as a function of incident pulse energy. The quantum efficiency in the linear regime is about 0.001, and the ultimate voltage is about 14 mV .

place of silicon, it appears possible to fabricate moderately efficient photodetectors with detection bandwidths in excess of 300 GHz.

This research was supported by the Newport Research Company and the Air Force Office of Scientific Research, with earlier support from the Lawrence Livermore National Laboratory. The author also wishes to thank S. J. B. Yoo for assistance in the fabrication of the detectors and in early impulse response measurements.

⁵S. Y. Wang and D. M. Bloom, *Electron. Lett.* **19**, 554 (1983).

⁶D. H. Auston, *Appl. Phys. Lett.* **26**, 101 (1975).

⁷R. B. Hammond, N. G. Paulter, R. S. Wagner, and T. E. Springer, *Appl. Phys. Lett.* **44**, 620 (1984).

⁸M. B. Ketchen, D. Grischkowsky, T. C. Chen, C.-C. Chi, I. N. Duling III, N. J. Halas, J.-M. Halbout, J. A. Kash, and G. P. Li, *Appl. Phys. Lett.* **48**, 751 (1986).

⁹R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits* (Wiley, New York, 1977), p. 31.

¹⁰K. Seeger, *Semiconductor Physics* (Springer, New York, 1973), pp. 155-157.

A Fast Diffusion-Driven Photodetector: Theory and Experiment

Adnah Kostenbauder

S. J. B. Yoo

A. E. Siegman

**Reprinted from
IEEE JOURNAL OF QUANTUM ELECTRONICS
Vol. 24, No. 2, February 1988**

A Fast Diffusion-Driven Photodetector: Theory and Experiment

ADNAH KOSTENBAUDER, S. J. B. YOO, AND A. E. SIEGMAN

Abstract—We have demonstrated a novel high-speed photodetector. Prototype silicon devices have detection bandwidths of 5 GHz and are limited by neither the saturation-velocity transit time nor the carrier lifetime. These detectors employ a combination of the Dember effect and the ease of generating fine photocarrier gratings, and appear to scale to bandwidths of several hundred gigahertz.

I. INTRODUCTION

PRESENT high-speed photodetectors can be divided into two classes. The first group uses small geometries to reduce the saturation-velocity transit time and circuit parasitics to acceptable levels. Devices of this type have obtained 3 dB bandwidths of 100 GHz with inductive peaking and 70 GHz without [1]. An important aspect of these devices is their high quantum efficiency. The second group of detectors relies upon shortened carrier lifetimes to rapidly terminate the photocurrent [2]–[4]. This technique has resulted in 3 dB bandwidths of about 300 GHz [4], but the quantum efficiencies of these devices are very low. We describe below a new type of high-speed detector [5] which is based upon the rapid decay of photocarrier gratings and the associated Dember-effect potentials. These devices appear to be capable of very high-speed operation, while maintaining reasonable quantum efficiencies.

II. THEORY

We begin by describing how a highly idealized diffusion-driven photodetector works. In particular, assume that a thin homogeneous slab of semiconductor is illuminated through both faces with the same short pulse of light. The position and thickness of the slab are chosen such that a null of the optical interference pattern occurs at one face, while a peak occurs at the other. The slab is assumed to be devoid of contacts and their associated junction fields. (It might, for example, be radiating its output into a nearby waveguide.) It is also assumed to be free of surface states and resultant band bending. Finally, we assume thermalized carrier distributions and no recombination. Although the idealizations made above re-

move most real-world complications, they are useful in that they allow an analytic development of the theory, and also aid the development of an intuitive picture of how the device works, which is applicable even when many of these idealizations are not made.

Under the assumptions made above, the dynamic equations for the carrier densities [6] and the electric field are

$$\begin{aligned}\vec{J}_n &= -D_n \nabla N - \mu_n \vec{\mathcal{E}} N \\ \vec{J}_p &= -D_p \nabla P + \mu_p \vec{\mathcal{E}} P \\ \frac{\partial N}{\partial t} &= -\nabla \cdot \vec{J}_n \\ \frac{\partial P}{\partial t} &= -\nabla \cdot \vec{J}_p \\ \nabla \cdot \vec{\mathcal{E}} &= \frac{q}{\epsilon} (P - N + \text{net doping})\end{aligned}\quad (1)$$

where N and P are the electron and hole carrier densities, respectively, \vec{J}_n and \vec{J}_p are the electron and hole fluxes, and $\vec{\mathcal{E}}$ is the time- and space-varying electric field. Under the conditions described in the previous paragraph, the boundary conditions are

$$\begin{aligned}\vec{J}_n(0, t) &= 0 = \vec{J}_n(L, t) \\ \vec{J}_p(0, t) &= 0 = \vec{J}_p(L, t) \\ \vec{\mathcal{E}}(0, t) &= 0 = \vec{\mathcal{E}}(L, t)\end{aligned}\quad (2)$$

where 0 and L are the surfaces of the device. The carrier densities just after the arrival of an optical impulse (taken to be at $t = 0$) are

$$\begin{aligned}N(x, 0) &= N_0 + Q \cos kx \\ P(x, 0) &= P_0 + Q \cos kx.\end{aligned}\quad (3)$$

Defining $n = N - N_0$ and $p = P - P_0$, combining the various parts of (1), and retaining only the terms linear in n and p yields

$$\begin{aligned}\frac{\partial n}{\partial t} &= D_n \nabla^2 n + \frac{\mu_n q N_0}{\epsilon} (p - n) \\ \frac{\partial p}{\partial t} &= D_p \nabla^2 p + \frac{\mu_p q P_0}{\epsilon} (n - p).\end{aligned}\quad (4)$$

Manuscript received May 29, 1987; revised August 4, 1987. This work was supported by the U.S. Air Force Office of Scientific Research and the Lawrence Livermore National Laboratory. The work of A. Kostenbauder was supported by a Newport Research Award.

The authors are with the Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305.

IEEE Log Number 8717794.

For convenience, we identify $\gamma_n \equiv \mu_n q N_0 / \epsilon$ and $\gamma_p \equiv \mu_p q P_0 / \epsilon$ as the dielectric relaxation rates associated with the two carrier types. The solution of (4), subject to the conditions (2) and (3), is

$$\begin{aligned} n &= \frac{Q}{2\beta} [(\beta + \alpha + D_n k^2) \exp [(\alpha - \beta)t] \\ &\quad + (\beta - \alpha - D_n k^2) \exp [(\alpha + \beta)t]] \cos kx \\ p &= \frac{Q}{2\beta} [(\beta + \alpha + D_p k^2) \exp [(\alpha - \beta)t] \\ &\quad + (\beta - \alpha - D_p k^2) \exp [(\alpha + \beta)t]] \cos kx \end{aligned} \quad (5)$$

where

$$\alpha = -[(D_n + D_p) k^2 + \gamma_n + \gamma_p]/2$$

and

$$\beta = (1/2) \sqrt{[(D_n - D_p) k^2 + \gamma_n - \gamma_p]^2 + 4\gamma_n \gamma_p}.$$

The corresponding electric potential within the slab is given by

$$\begin{aligned} V(x, t) &= \frac{Qq}{2\beta\epsilon} (D_n - D_p) (\exp [(\alpha + \beta)t] \\ &\quad - \exp [(\alpha - \beta)t]) \cos kx. \end{aligned} \quad (6)$$

The potential difference between the faces of the slab is identical to (6), but with $\cos kx$ replaced by 2.

Two special cases of (6) are of particular interest. First is the case of low-level injection in material doped so that γ_p far exceeds $D_n k^2$, $D_p k^2$, and γ_n (i.e., heavily p-type material). In this case, (6) reduces to

$$\begin{aligned} V &= \frac{Qq(D_n - D_p)}{\gamma_p \epsilon} (\exp(-D_n k^2 t) - e^{-\gamma_p t}) \cos kx \\ &\approx \frac{Qq(D_n - D_p)}{\gamma_p \epsilon} \exp(-D_n k^2 t) \cos kx \end{aligned} \quad (7)$$

with an obvious analog for heavily n-doped material. Second is the case where the injected carrier densities are very large compared to the equilibrium densities, and γ_n and γ_p are both large compared to the diffusion rates. Then, (6) reduces to

$$\begin{aligned} V &= \frac{Qq(D_n - D_p)}{(\gamma_n + \gamma_p)\epsilon} (\exp(-D_a k^2 t) \\ &\quad - \exp[-(\gamma_n + \gamma_p)t]) \cos kx \\ &\approx \frac{Qq(D_n - D_p)}{(\gamma_n + \gamma_p)\epsilon} \exp(-D_a k^2 t) \cos kx \end{aligned} \quad (8)$$

where $D_a = 2D_n D_p / (D_n + D_p)$ is the ambipolar diffusion coefficient. The second exponentials in (7) and (8) can usually be dropped as their only effect is a very fast rise time.

The physical source of the potentials in (6), (7), and (8) is the difference in the speed of electron and hole diffu-

sion. When the carrier grating is written into the slab, there is no net space charge anywhere, and the electrons begin to race out into the troughs of the distribution more quickly than do the holes. Under typical conditions, a very small fractional separation of the carriers results in an electric field (the Dember effect [6]) that prevents further separation. The time required to create this field is described by the second exponentials in (7) and (8). Once this field has been created, the two carrier gratings decay in lock step—as if both carriers had the same diffusion coefficient. In reality, however, the diffusion constants are different, and there is an electric field within the device which is holding the carriers together and giving a photovoltaic signal.

We now begin to remove the idealizations introduced at the beginning of this section. First we want to place (symmetrical) contacts on our slab so that the generated signal can be carried away in a stripline. The introduction of contacts prevents analytic solutions to (1) and might be expected to destroy the orderly grating decay picture presented above. However, computer simulations of (1), using silicon's material parameters, indicate that the temporal dependence of the potential in (7) is not greatly disturbed in the presence of weak junction fields. Specifically, the time constant changes by less than a factor of two for fields less than a thermal voltage per grating cycle. This is due to the fact that the generation of a potential difference between the two surfaces requires a symmetry-breaking element within the device, and the device is symmetrical except for the carrier grating. Moreover, this grating will decay by diffusion while drifting due to the junction fields. Although the temporal dependence of the photovoltage is not substantially altered, its magnitude can be substantially altered due to drift of the grating, even for small fields.

Given the sensitivity of the photovoltage's magnitude to junction fields, it is possible to obtain only a zeroth estimate of the quantum efficiency (into a short circuit load) from (6)

$$\eta = \frac{2(\gamma_n + \gamma_p) |D_n - D_p|}{l^2 ((D_n k^2 + \gamma_n)(D_p k^2 + \gamma_p) - \gamma_n \gamma_p)}. \quad (9)$$

This reduces to

$$\eta = \frac{2|D_n - D_p|}{D_n} (kl)^{-2} \quad (10)$$

in the low-level injection case and to

$$\eta = \left| \frac{D_n}{D_p} - \frac{D_p}{D_n} \right| (kl)^{-2} \quad (11)$$

in the ambipolar case.

Recombination effects are usually negligible in describing the operation of a diffusion-driven detector; however, at large fringe spacings, they do contribute to the decay of the carrier gratings and may be accounted for by adding the recombination rate to both of the rates in the expo-

nentials in (6) with corresponding minor changes in (7) through (11).

Finally, the smallest fringe spacing, and hence the fastest response time, is obtained when the two optical beams are counterpropagating so that $k = 4\pi n/\lambda$ where λ is the vacuum wavelength and n is the semiconductor's index of refraction. For visible wavelengths, this corresponds to $k \approx 5 \times 10^5 \text{ cm}^{-1}$ which, combined with a diffusion constant of $20 \text{ cm}^2/\text{s}$, gives the first exponentials in (7) and (8) time constants of just under 200 fs.

Clearly, this limiting case falls outside the regime of equilibrium carrier diffusion, and the detector response must be calculated taking into account phase space considerations. As a specific example of this regime, assume that the detector is fabricated from lightly doped GaAs and is hit with a brief optical pulse centered at 730 nm. By energy and momentum conservation considerations, this creates electrons 240 meV above the band edge (i.e., the phase space electron density is $(1 + \cos kx) \delta(\text{kinetic energy} - 240 \text{ meV})$ and holes with kinetic energies of 24 meV. Ferry has calculated [8] that electrons of this energy injected into $N = 10^{17}/\text{cm}^3$ GaAs lose about 20 percent of their velocity in traveling 60 nm under small electric fields ($|E| < 1 \text{ kV/cm}$) and this loss should be even lower for lightly doped material. Since the electrons need only travel one peak-to-trough distance (60 nm) to wash out the grating, a reasonable estimate of the electron density versus time may be obtained by treating the electrons as if their velocities are conserved. The result is a spatial density of injected electrons proportional to $1 + \cos(kz) \sin(kvt)/(kvt)$ where v is the electron group velocity at 240 meV ($\approx 10^8 \text{ cm/s}$). The grating part of this decays to $1/e$ at $t \approx 50 \text{ fs}$. We stress that this time is much less than the measured carrier-phonon collision time [9] and that, as a consequence, the assumption of velocity conservation is very reasonable. On the other hand, the holes are injected at thermal energies and (from the Drude model) have a scattering time about half that of the electrons. Thus, it is reasonable to treat the holes as if in equilibrium. The hole density is thus proportional to $1 + \cos(kx) \exp(-D_p k^2 t)$. Combining these yields an impulse response proportional to $\exp(-D_p k^2 t) - \sin(kvt)/kvt$. The FWHM of this is 210 fs, which corresponds to a bandwidth of many hundreds of GHz.

III. EXPERIMENT

We have fabricated two sets of prototype silicon diffusion-driven photodetectors. The first was made by heavily doping both surfaces of $100 \mu\text{m}$ thick high-resistivity ($N = 2 \times 10^{13} \text{ cm}^{-3}$) silicon wafers, and depositing about $1 \mu\text{m}$ of aluminum over the surfaces. These wafers were then immersed in LN_2 and cleaved into rectangles from 1 to 2 mm on a side. Acceptable pieces were bonded to the back side of SMA bulkhead connectors with one edge protruding beyond the flange, as shown in Fig. 1(a). The other aluminum surface was ultrasonically bonded to a fine wire running to the center conductor. The lead in-

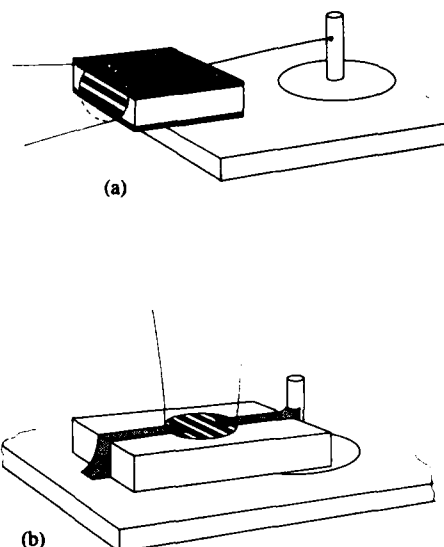


Fig. 1. (a) The geometry of a diffusion-driven detector of the first generation. (b) Represents the second generation gap in a stripline configuration.

ductance in this arrangement was about 1 nH, while the capacitance between the aluminum contacts was a few pF.

The second set of devices was fabricated as a $25 \mu\text{m}$ gap in a $300 \mu\text{m}$ wide aluminum line on one side of $380 \mu\text{m}$ thick silicon, and is illustrated in Fig. 1 (b). The silicon used here had the same resistivity as in the first devices, and the aluminum was $0.2 \mu\text{m}$ thick. The ratio of linewidth to wafer thickness was chosen to give a 50Ω line, with parasitics that are negligible at all frequencies measured ($< 10 \text{ GHz}$).

Both sets of samples were excited with a synchronously pumped mode-locked dye laser operating near 590 nm. The pulses had a duration of less than 5 ps and energies of about 10 nJ. The first devices were excited at a repetition rate of 150 kHz, while the second set was operated at 4 MHz. In order to produce a variety of fringe spacings on the detectors, the laser beam was expanded and transmitted through a Ronchi ruling which was imaged onto the active surface of the detector. We produced a wide range of k vectors on the detector by using various rulings and several different microscope objectives to image the ruling. The output of the detector was directly connected to a Tektronix S-4 sampling head.

The intensity pattern produced by this apparatus is not a simple sinusoid, but rather a sinusoidally modulated Gaussian. This pattern can be decomposed as a sum of sinusoids which obey the boundary conditions (2). In this decomposition, the only sinusoids which have substantial amplitude are near $k = 0$ and the value of k expected from the imaging of the ruling. This results in a photocurrent which has an initial fast component whose sign depends on the transverse position of the ruling, and a slow component whose sign depends on where the detector is in the spot. Fortunately, the slow component nulls when the detector is centered in the spot, leaving only the fast grating-related component.

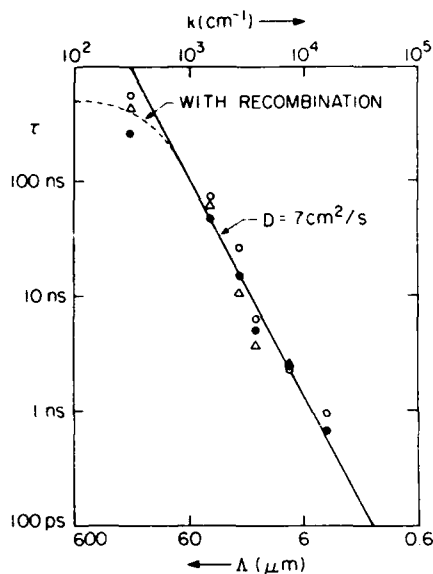


Fig. 2. Time constants in three first generation detectors as a function of fringe spacing Λ ($\Lambda = 2\pi/k$).

Signal levels ranged from 2 to 5 mV for the first class of detector, and were from 5 to 10 mV for the second group. Oscilloscope photographs were digitized and fit to single exponentials. Figs. 2 and 3 show the dependence of the time constant on the grating space in the first and second classes of devices, respectively. The ambipolar diffusion constant inferred from Fig. 1 is $7 \pm 2 \text{ cm}^2/\text{s}$, while that determined from Fig. 2 is $8.6 \pm 2 \text{ cm}^2/\text{s}$. Given the uncertainties introduced by junction fields, these are in remarkable agreement with silicon's surface ambipolar diffusion constant. As shown in Fig. 2, a correction for recombination needs to be made at the largest grating spacings tested, suggesting a surface recombination velocity of about 1000 cm/s, as the bulk lifetime in this silicon was 400 μs before the fabrication of the contacts. This nondiffusive time constant may also be due to junction fields, as the small potential per grating cycle approximation is badly violated here.

The shortest response times measured were 350 and 40 ps for devices of the first and second geometries, respectively. These are substantially below the electron saturation-velocity transit times which were 1 ns and 250 ps, respectively. Since the carrier lifetime in both of these devices is at least several hundred nanoseconds, we have also verified that this device is not governed by the carrier lifetime limit applicable to photoconductors.

As can be seen in (11), the quantum efficiency scales as $(kl)^{-2}$ in the ambipolar regime. The k^{-2} part of this dependence appears to be well obeyed in the devices of both groups, although the proportionality constant is not well determined due to geometrical losses. Also, the quantum efficiency for a fixed value of k appears to be substantially higher for the smaller devices, although again geometrical effects prevent precise comparison.

Finally, we have measured the charge driven through the 50 Ω load by one of the devices of the first type op-

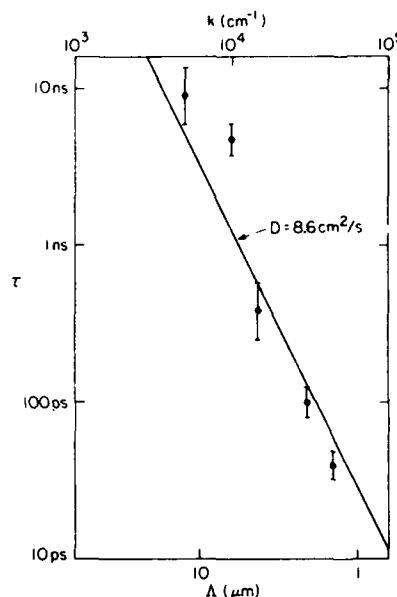


Fig. 3. Response time of second generation detectors versus k . The smallest grating period used was 1.4 μm .

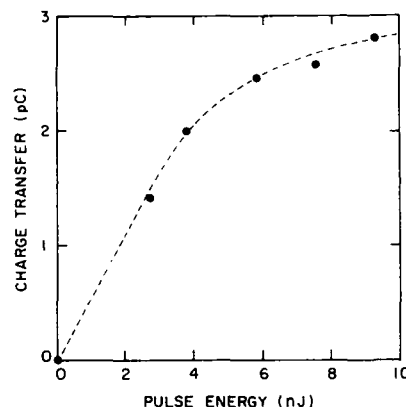


Fig. 4. Saturation of a first generation detector with a fringe spacing of 40 μm . The quantum efficiency in the linear regime is about 10^{-3} without correction for geometrical losses.

erating with a fringe spacing of 40 μm as a function of input pulse energy. These results are presented in Fig. 4 and show a linear response at low energies ($\eta = 0.001$ without correction for geometrical factors) and a saturation at high fluences. We believe that at high fluxes, the internal resistance of the device becomes comparable to the 50 Ω load, resulting in much of the generated current flowing back through the device rather than out into the coaxial line.

IV. CONCLUSIONS

We have demonstrated a novel type of photodetector which is based upon the combination of the Dember effect and the ease of optically generating fine carrier gratings. Prototype device response times are limited neither by the saturation-velocity transit time nor by the carrier lifetime. Present devices have detection bandwidths of 5 GHz, and scaling to detectors which use anti-parallel optical beams

to write the carrier gratings should achieve bandwidths of several hundred gigahertz.

REFERENCES

- [1] S. Y. Yang and D. M. Bloom, "100 GHz bandwidth planar GaAs Schottky photodiode," *Electron. Lett.*, vol. 19, pp. 554-555, 1983.
- [2] D. H. Auston, "Picosecond optoelectronic switching and gating in silicon," *Appl. Phys. Lett.*, vol. 26, pp. 101-103, 1975.
- [3] C. H. Lee, "Picosecond optoelectronic switching in GaAs," *Appl. Phys. Lett.*, vol. 30, pp. 84-86, 1977.
- [4] M. B. Ketchen, D. Grischowsky, T. C. Chen, C.-C. Chi, I. N. Duling, III, N. J. Halas, J.-M. Halbout, J. A. Kash, and G. P. Li, "Generation of subpicosecond electrical pulses on coplanar transmission lines," *Appl. Phys. Lett.*, vol. 48, pp. 751-753, 1986.
- [5] A. G. Kostenbauder, "A high-speed diffusion-driven photodetector," submitted to *Appl. Phys. Lett.*
- [6] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981, pp. 50-51.
- [7] K. Seeger, *Semiconductor Physics*. New York: Springer, 1973, pp. 155-157.
- [8] P. Lugli and D. K. Ferry, "Carrier-carrier interaction and picosecond phenomena in polar semiconductors," in *Picosecond Electronics and Optoelectronics*, G. A. Mourou, D. M. Bloom, and C.-H. Lee, Eds. New York: Springer-Verlag, 1985, pp. 83-86.
- [9] J. A. Kash, J. C. Tsang, and J. M. Hvam, "Subpicosecond Raman spectroscopy of electron-LO phonon dynamics in GaAs," in *Picosecond Electronics and Optoelectronics*, G. A. Mourou, D. M. Bloom, and C.-H. Lee, Eds. New York: Springer-Verlag, 1985, pp. 87-90.

Adnah Kostenbauder, photograph and biography not available at the time of publication.

S. J. B. Yoo, photograph and biography not available at the time of publication.

A. E. Siegman, photograph and biography not available at the time of publication.

BIPOLAR OPTICAL MODULATION AND DEMODULATION USING A DUAL-MODE FIBER AND A FAST DIFFUSION-DRIVEN PHOTODETECTOR

Paul Wysocki, A. G. Kostenbauder,
B. Y. Kim, and A. E. Siegman
Edward L. Ginzton Laboratory
Stanford University, Stanford CA 94305

ABSTRACT

We demonstrate a bipolar fiber-optic signal processing and photodetection concept which combines a dual-mode optical fiber with a fast diffusion-driven photodetector.

A dual-mode fiber with an elliptical core of the proper dimensions can propagate a symmetric lowest-order LP₀₁ mode, or an antisymmetric LP₁₁ mode, but not any higher-order modes [1]. Suppose a coherent superposition of these two modes is excited in the fiber with comparable amplitudes but an adjustable phase difference between the two modes. The transverse intensity profile across the end of the fiber, measured along the major axis of the ellipse, will then have one or the other of the two forms shown in Figure 1 if the two modes are excited either in phase or 180 degrees out of phase with each other at the output end. Intermediate phase angles will produce intermediate intensity profiles. In physical terms, the intensity distribution can be skewed or shifted back and forth from one side of the fiber to the other by varying the relative optical phase difference between the two modes.

One of us (A.G.K.) has recently demonstrated a novel, very fast, diffusion-driven photodetector [2,3] which will respond in a bipolar fashion to precisely the

sort of spatially asymmetric optical intensity distributions shown in Figure 1. That is, each of these intensity distributions can be viewed as being more or less a single period of an off-center intensity grating pattern, with a transverse shift in position of roughly one grating period between the two cases. The diffusion-driven photodetector, if properly aligned, will respond to these skewed intensity patterns with a positive photovoltaic output for a light pattern which is displaced to one side, and a negative photovoltaic output for a light pattern displaced to the other side. A symmetric or centered intensity pattern, such as is produced by either of the fiber modes alone, will produce no response from this photodetector. Note that the detector is purely photovoltaic in character, i.e., it requires no dc bias voltage. Detectors of this type have been found to exhibit pulse responses to properly skewed optical signals with response times from nanoseconds down to picoseconds [3], and sub-picosecond responses have been observed in unpublished measurements.

Figure 2 shows an elementary bipolar optical signal processing or photodetection system which can be implemented by combining these dual-mode fiber and diffusion-driven photodetector concepts. Coherently related LP_{01} and LP_{11} mode signals are assumed to be excited in the dual-mode fiber. The output light intensity pattern coming from the fiber is then imaged, properly centered, to more or less fill the gap of a diffusion-driven photodetector of the type described earlier [3]. The dual-mode fiber might be excited, for example, with one transverse mode already present in the fiber and the other mode injected through a suitable mode-selective fiber-optic coupler [4]. One or both of the light signals is assumed to be pulsed.

Neither input signal by itself will then produce an output from the photodetector, assuming the fiber and photodetector are properly centered. When both signals are present and overlapping in time, however, output pulses will be detected with one electrical polarity if the two signals are optically in phase at the fiber output end, or with opposite polarity if the two signals are 180 degrees out of phase, or with any amplitude in between if the two signals have intermediate phases. A

change in the optical phase of either signal—such as might be produced either by an electrooptic modulator or through light-by-light modulation [5]—can thus be used to vary the amplitude and polarity of the pulse detection of the other signal. Only one of the two signals in fact needs to be a pulse train; the other could equally well be a coherently related cw signal. A variety of optical signal processing systems could be implemented using this general approach.

As an elementary demonstration of this concept we performed the simple experiment shown in Figure 3. Optical pulses approximately 200 ps in duration from a mode-locked and frequency-doubled Nd:YAG laser were coupled into a 10 meter length of dual-mode fiber with the input beam slightly offset so as to excite approximately equal amplitudes of the LP_{01} and LP_{11} modes in the fiber. The dual-mode fiber had a lowest-order mode size of approximately 2 microns \times 4 microns and was provided by the Polaroid Corporation. The output end of the fiber was imaged at between 5 and 10 times magnification onto the 25 micron gap in a silicon gap-type diffusion-driven photodetector as described by Kostenbauder [3]. The relative phase angle between the two modes at the output end of the fiber could then be scanned either by mechanically stretching the fiber [6] or, at higher modulation speeds, by wrapping the fiber tightly around a cylindrical piezoacoustic element which was electrically driven at its lowest resonance frequency of 25 kHz. Mechanical stress in the wrapped fiber then led to a periodic relative phase modulation between the two modes. The output from the unbiased diffusion-driven photodetector was connected through a broadband amplifier with an upper cutoff frequency of 300 MHz to the input of a fast oscilloscope, giving an overall response time of approximately 1 ns.

Examples of the resulting photodetector output are shown in Figure 4. Figures 4(a) and 4(b) show sections of the 80 MHz mode-locked pulse train detected with positive and negative polarity as a result of opposite values of the relative phase shift between the two modes. Figure 4(c) shows the positive and negative modulation of the detection envelope for the 80 MHz optical pulse train over several

cycles of the 25 kHz phase modulation frequency. The oscilloscope trace has been retouched to reduce baseline smearing and distortion caused by the low duty cycle of the laser pulses and by inadequate low-pass response of the amplifier following the photodetector. Careful positioning of the diffusion-driven photodetector was required to obtain a symmetrically balanced output as shown.

Various permutations of the ideas described in this note can permit different kinds of coherent and incoherent optical signal processing and optical control of pulsed photodetection. If the end of a typical dual-mode fiber, with an intensity pattern width of approximately 4 microns is imaged onto a suitable photodetector gap with a 2 times demagnification, a photodetector response time of approximately 400 ps can be obtained. Various forms of integrated dual-mode fiber and photodetector combinations to obtain the bipolar photodetection without active alignment requirements can also be envisioned.

This work was supported by the Air Force Office of Scientific Research. The authors thank C. C. Pohalski and K. A. Fesler for assistance with the experiments. and P. Wysocki thanks the National Science Foundation for fellowship support.

REFERENCES

1. B. Y. Kim, J. N. Blake, S. Y. Huang, and H. J. Shaw, "Use of highly elliptical core fibers for two-mode fiber devices," *Opt. Lett.* **12**, 729 (1987).
2. Adnah G. Kostenbauder, "A high speed diffusion-driven photodetector," *Appl. Phys. Lett.* **51**, 1129-1131 (April 1987).
3. Adnah Kostenbauder, S.J.B. Yoo, and A.E. Siegman, "A fast diffusion-driven photodetector: Theory and experiment," *IEEE J. Quantum Electron.* **24**, 240-244 (February 1988).
4. W. V. Sorin, B. Y. Kim, and H. J. Shaw, "Highly selective evanescent modal filter for two-mode optical fibers," *Opt. Lett.* **11**, 581 (1986).

5. H. G. Park, C. C. Pohalski, and B. Y. Kim, "Optical Kerr switch using elliptical-core two-mode fiber," *Opt. Lett.* **13**, 776-778 (September 1988).

6. J. N. Blake, S. Y. Huang, B. Y. Kim, and H. J. Shaw, "Strain effects on highly elliptical core two-mode fibers," *Opt. Lett.* **12**, 732 (1987).

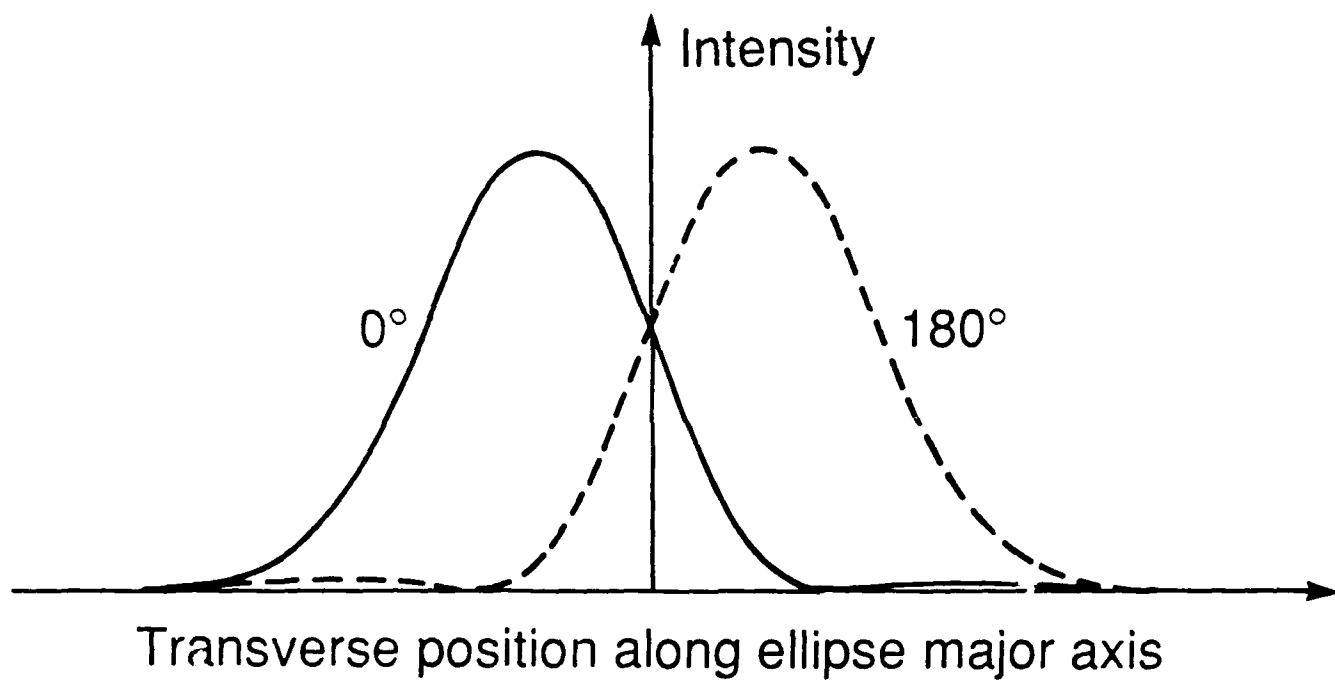
CAPTIONS FOR FIGURES

Figure 1. Transverse intensity profile in the x direction across the end of a two-mode fiber propagating the LP_{01} and LP_{11} modes when the two modes are (a) optically in phase, and (b) 180 degrees out of phase.

Figure 2. An optical signal processing system in which the two lowest-order modes are combined in a single fiber with variable phase difference and then optically coupled to a diffusion-driven photodetector.

Figure 3. Experimental system for demonstrating the bipolar dual-mode photodetection concept. The two lowest-order propagation modes are propagated in a dual-mode fiber tightly wound about a piezoelectric cylinder which creates mechanical strain and thus relative phase modulation between the two modes at an audio frequency rate.

Figure 4. Bipolar detection of mode-locked pulse trains resulting from the low-frequency modulation of the relative phase between LP_{01} and LP_{11} modes. (a) and (b) Individual mode-locked pulses detected with positive and negative polarity at the positive and negative peaks of the audio-frequency phase modulation envelope (time scale = 20 ns/division). (c) Envelope of the detected mode-locked pulse train over several periods of the 25 kHz phase modulation (time scale = 20 μ s/division).



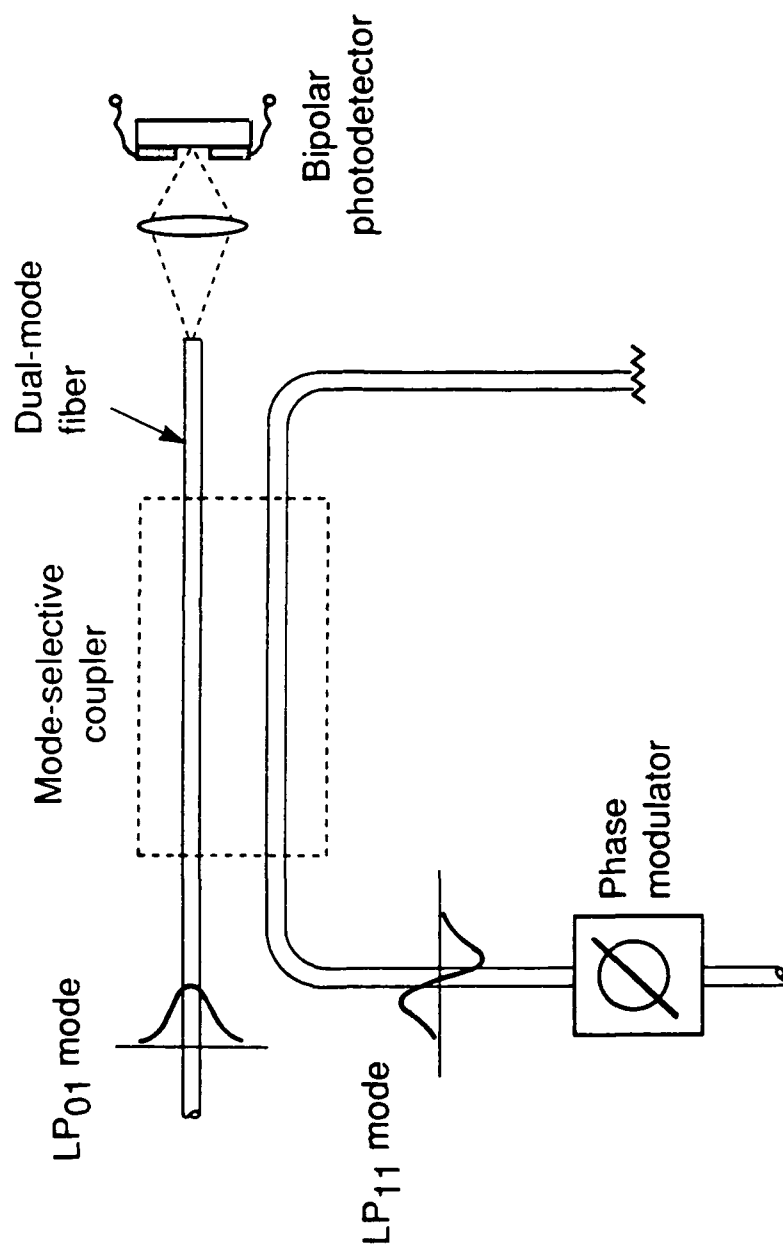


FIG. 2

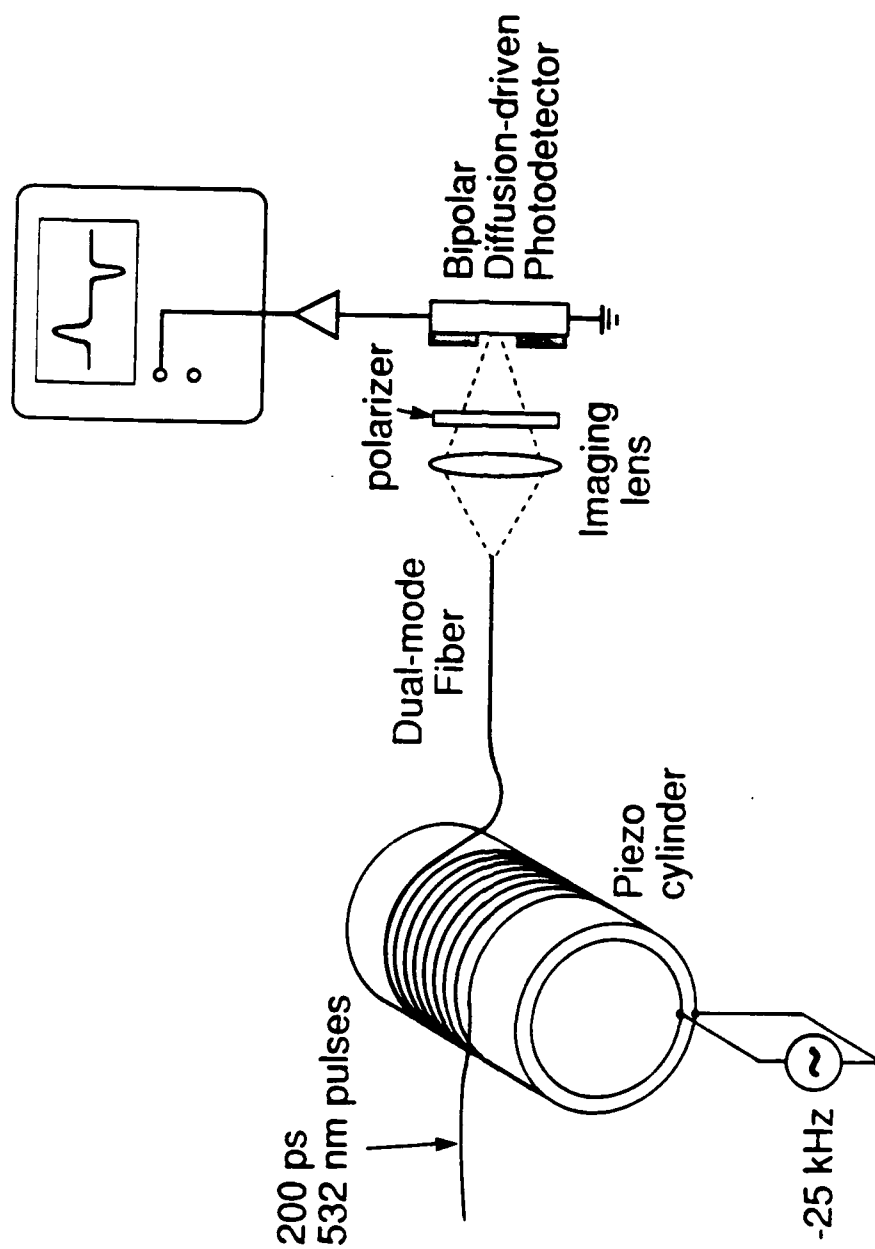
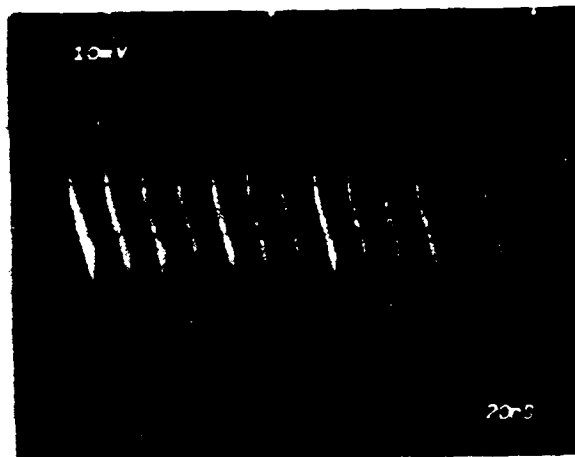


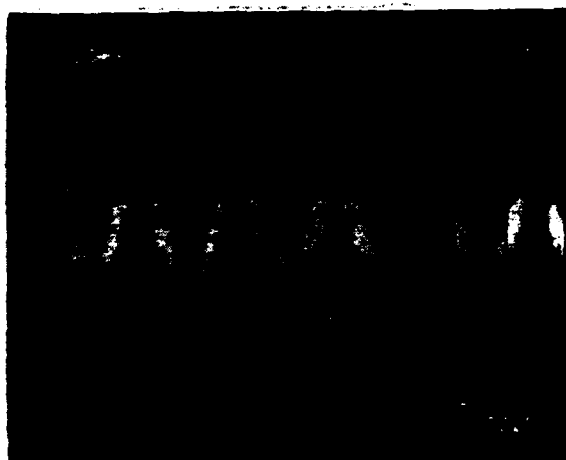
Fig. 3



(a)



(b)



(c)

Excess spontaneous emission in non-Hermitian optical systems. I. Laser amplifiers

A. E. Siegman

*Max-Planck Institut für Quantenoptik, D-8046 Garching, West Germany
and E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305*

(Received 15 August 1988)

Petermann first predicted in 1979 the existence of an excess-spontaneous-emission factor in gain-guided semiconductor lasers. We show that an excess spontaneous emission of this type, and also a correlation between the spontaneous emission into different cavity modes, will in fact be present in all open-sided laser resonators or optical lens guides. These properties arise from the non-self-adjoint or non-power-orthogonal nature of the optical resonator modes. The spontaneous-emission rate is only slightly enhanced in stable-resonator or index-guided structures, but can become very much larger than normal in gain-guided or geometrically unstable structures. Optical resonators or lens guides that have an excess noise emission necessarily also exhibit an "excess initial-mode excitation factor" for externally injected signals. As a result, the excess spontaneous emission can be balanced out and the usual quantum-noise limit recovered in laser amplifiers and in injection-seeded laser oscillators, but not in free-running laser oscillators.

I. INTRODUCTION

The conventional second-quantized analysis of spontaneous emission in a laser cavity or in any other system involving interaction between atoms and electromagnetic radiation leads to a general principle that the rate of spontaneous emission from a collection of atoms into any individual resonant-cavity or transmission-line mode will always be exactly equal to the downward stimulated transition rate that would be produced in the same atoms by a signal energy of one extra photon in the same electromagnetic mode. This principle can be used to derive fundamental conclusions relating to thermal equilibrium, blackbody radiation density, Johnson-Nyquist or thermal noise in lossy linear systems, laser-amplifier noise figure, and quantum-noise fluctuations in laser oscillators.

Petermann first predicted in 1979 a so-called "excess-spontaneous-emission factor" or excess-noise factor with a value greater than one extra photon per mode in gain-guided semiconductor lasers.¹ This prediction was initially controversial, since spontaneous emission at a rate corresponding to more than one extra photon per mode would seem to violate the accepted principle of quantum-noise theory mentioned above. Moreover, application of the same argument to passive or loss-guided systems appears to predict an excess thermal noise in such systems, in apparent violation of elementary thermodynamics.

These difficulties were resolved, at least for loss-guided systems, in an excellent analysis by Haus and Kawakami,² who pointed out that such gain or loss-guided systems also exhibit correlations between the noise emission into different propagating modes. This is in contrast to the familiar situation with power-orthogonal normal modes, where the spontaneous emission into each mode is uncorrelated with the emission into any other mode. Haus and Kawakami showed that the correlations between emission into different modes in

the loss-guided case are just sufficient to recover the usual blackbody radiation or Johnson-Nyquist thermal-noise results in these loss-guided systems. Several other derivations of Petermann's excess-noise factor have also been presented by other authors;³⁻¹³ most of these are reviewed in Haus and Kawakami. An earlier discussion of spontaneous emission in open resonators and its application to laser amplifiers and oscillators, in much the same spirit as the present paper, has also been given by Henry.¹⁴

In this paper we demonstrate that an excess spontaneous emission rate per mode and also noise correlations between the emission into different modes are, in fact, to be expected in *all* open-sided laser resonators, optical lens guides, and similar optical structures, exactly as predicted by Petermann and by Haus and Kawakami for the gain-guided case. These excess-noise properties have nothing necessarily to do with gain or loss guiding, or with wave-front curvature. They arise entirely from the non-Hermitian and hence non-power-orthogonal or biorthogonal nature of the transverse eigenmodes in such structures.^{15,16} These excess-noise emission and mode-correlation effects remain small (normalized values close to unity or zero, respectively) in conventional stable optical resonators or index-guided systems. They become of significant magnitude, however, in gain-guided or loss-guided systems, or in unstable-resonator or unstable-lens-guide systems having significant geometric magnification per pass. In such systems, the excess emission factor can become as large as hundreds to thousands of times above the usual value, even at moderate Fresnel numbers for the unstable systems.

The crucial step in the analysis in this paper comes in using the real Fox and Li transverse eigenmodes of the optical system¹⁵ as the basis set for expanding the fields in the structure, rather than assuming an idealized set of power-orthogonal normal modes as is usually (and incorrectly) done in most laser analyses. The biorthogonal

rather than power-orthogonal nature of the real cavity modes then turns out to be responsible for the excess-noise and correlated-emission properties. The approach in this paper is very similar to the Haus-Kawakami approach, but the results now apply to a much broader class of general, open-sided, non-loss-guided optical structures and the analogous periodic lens guides.

We also demonstrate that any such systems having significant excess-noise emission will necessarily also exhibit a corresponding "excess initial mode excitation" in their response to externally injected signals. By sending in a properly shaped "adjoint-coupled" external signal, for example, it is possible to excite such systems with more initial power or energy in any given transverse eigenmode of the system than is present in the injected external signal itself. As a result, in a laser amplifier or an injection-seeded laser oscillator with optimum signal injection, one can always recover the minimum quantum-limited noise performance predicted by conventional theories, i.e., one equivalent input noise photon per resolution time for the lowest-loss mode, despite the excess spontaneous emission with the laser amplifier.

The situation is less satisfactory for free-running unstable-resonator laser oscillators, however. In the second part of this paper we will show that a laser oscillator can always be expected to exhibit quantum noise effects or Schawlow-Townes noise fluctuations which are larger than the usually stated values by just the excess-spontaneous-emission ratio or excess-noise factor calculated in this paper. Fortunately, this excess-noise factor is close to unity for lasers using ordinary stable laser cavities, or using pure index guiding. For gain-guided lasers, however, or for either hard-edged or variable-reflectivity unstable resonators, the excess-noise enhancement can be as large as 100 to 1000 times for moderate values of geometrical magnification and Fresnel number.

II. NOISE ANALYSIS FOR PERIODIC LENS-GUIDE SYSTEMS

In this paper we will use an extended semiclassical analysis to derive the excess-noise properties for a laser

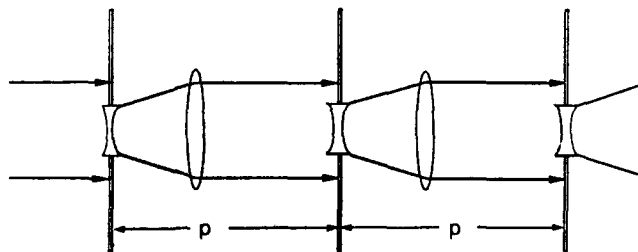


FIG. 1. Example of a periodic lens guide filled with absorbing or amplifying atomic medium, such as analyzed in this paper. The lens guide shown is a geometrically unstable confocal system to emphasize the characteristics of such systems. The absorbing apertures separating each period of the structure represent the diffractive loss out the sides or past the edges of the structure in each period. As such, they are taken to be perfectly absorbing but noise free or effectively at zero temperature.

amplifier produced by spontaneous emission from inverted laser atoms within the optical structure. It will be most convenient for this purpose to consider amplification down a cascaded periodic lens-guide system, as in Fig. 1, and to express all signal and noise quantities in the frequency domain. In a subsequent paper we will modify the analysis to apply to the time buildup of laser oscillation in a resonant laser cavity.

A. Analytical formulation

We wish to analyze, therefore, the propagation of waves traveling in the forward direction down a periodic optical waveguide or lens guide completely filled with an amplifying (or possibly absorbing) atomic medium, as illustrated in Fig. 1. This could, of course, also represent amplification through successive passes around an optical resonator containing the same atomic medium. We have made the example shown in Fig. 1 a geometrically unstable lens guide with a sizable magnification per period, to emphasize the characteristics of such systems. The analysis is intended to apply, however, to either stable or unstable systems, with the assumption that all such systems will have some finite-diameter aperture or mirror, and hence at least some amount of diffractive energy loss past the mirror edges or out the sides of the periodic structure.

We begin as usual with the scalar wave equation

$$\nabla^2 \mathcal{E}(\mathbf{r}, t) - \mu \sigma \frac{\partial \mathcal{E}(\mathbf{r}, t)}{\partial t} - \mu \epsilon \frac{\partial^2 \mathcal{E}(\mathbf{r}, t)}{\partial t^2} = \mu \frac{\partial^2 p_N(\mathbf{r}, t)}{\partial t^2}, \quad (1)$$

where $\mathcal{E}(\mathbf{r}, t)$ gives the real E field of the propagating wave in the medium as a function of space coordinates \mathbf{r} and time t . The conductivity σ (which is negative for an amplifying medium) represents the net linear stimulated emission or absorption by the atomic medium with which the E field interacts, and μ and ϵ are the magnetic and dielectric permeabilities of the medium. All these quantities are independent of position within the structure, i.e., there is no gain or index guiding (though this could easily be added, at the cost of some complexity in the analysis, and with no change in the overall results).

The polarization $p_N(\mathbf{r}, t)$ on the right-hand side of the equation is a random-noise polarization (dipole moment per unit volume) representing the spontaneous emission from the laser medium. The value of this noise polarization is derived from a heuristic argument in Appendix A.

For the amplifier case it is most convenient to assume a traveling-wave expansion in the form

$$\begin{aligned} \mathcal{E}(\mathbf{r}, t) &= \text{Re} \tilde{E}(\mathbf{r}, \omega) \exp[j(\omega t - \beta z)], \\ p_N(\mathbf{r}, t) &= \text{Re} \tilde{P}_N(\mathbf{r}, \omega) \exp[j(\omega t - \beta z)], \end{aligned} \quad (2)$$

where $\beta = \omega(\mu\epsilon)^{1/2}$. We omit writing an explicit Fourier integration over frequency ω since the system is linear, the noise signals at different frequencies are uncorrelated, and we will be examining the behavior only in a narrow bandwidth about any given carrier frequency ω . We can then make a slowly-varying-envelope approximation for

the phasor amplitude $\tilde{E}(\mathbf{r}, \omega)$, drop the explicit dependence on the frequency ω , and simplify the wave equation to the extended paraxial form

$$\nabla_T^2 \tilde{E} - 2j\beta \left[\frac{\partial \tilde{E}}{\partial z} - \alpha \tilde{E} \right] = -\omega^2 \mu \tilde{P}_N(\mathbf{r}), \quad (3)$$

where ∇_T^2 indicates the Laplacian operator with respect to the transverse coordinates x, y and $\alpha = -\eta_0 \sigma / 2$ is the gain coefficient in the atomic medium, with $\eta_0 = (\mu/\epsilon)^{1/2}$ being the characteristic impedance for the medium.

As shown in Appendix A, the phasor amplitude $\tilde{P}_N(\mathbf{r})$ for the noise polarization will have a δ -function correlation in space with an amplitude given by

$$\begin{aligned} \langle \tilde{P}_N(\mathbf{r}) \tilde{P}_N^*(\mathbf{r}') \rangle &= \frac{16\hbar\alpha B}{\omega\eta_0} \frac{N_2}{N_2 - N_1} \delta(\mathbf{r} - \mathbf{r}') \\ &= \frac{16\rho_I \hbar\alpha B}{\omega\eta_0} \delta(\mathbf{r} - \mathbf{r}'), \end{aligned} \quad (4)$$

where N_1 and N_2 are lower- and upper-level population densities, and B is the (Hertzian) bandwidth over which the noise power is measured. We also introduce the symbol ρ_I as a shorthand for the additional noise-emission factor $N_2/(N_2 - N_1)$ produced by incomplete inversion in the atomic medium. Equation (4) is essentially the same equation as has been used by other authors.^{1,2,14} Note that the mean-square amplitude of this spontaneous emission is directly proportional to the atomic gain coefficient α , since we assume the inversion N_2/N_1 is fixed and independent of position. We have written the gain coefficient and population ratios here as if the atomic medium were an inverted or amplifying medium. If the signs of the gain coefficient, population difference, and Boltzmann temperature are all inverted, however, the noise polarization magnitude will still remain positive, and all of the subsequent results will apply equally well to a lossy or absorbing atomic system.

B. Expansion in propagation eigenmodes

We then expand the total field in the optical system in terms of the "cold" propagating modes or transverse eigenmodes $\tilde{u}_n(\mathbf{s}, z)$ of the propagating structure¹⁵ in the form

$$\tilde{E}(\mathbf{r}) = \sum_n \tilde{c}_n(z) \tilde{u}_n(\mathbf{r}) = \sum_n \tilde{c}_n(z) \tilde{u}_n(\mathbf{s}, z), \quad (5)$$

where from here on we will use the notation $\mathbf{r} \equiv (\mathbf{s}, z)$ in order to distinguish the transverse coordinates $\mathbf{s} = (x, y)$ or $\mathbf{s} = (r, \theta)$ in the lens guide from the axial coordinate z . The functions $\tilde{u}_n(\mathbf{r}) \equiv \tilde{u}_n(\mathbf{s}, z)$ are intended to be the usual transverse eigenmodes, or "Fox and Li modes," of the optical resonator or lens guide without the amplifying medium. (For the case of two transverse dimensions the index n should really be a double set, i.e., $\tilde{u}_{nm}(\mathbf{s}, z)$, but we write only a single index for simplicity.)

These transverse eigenmodes \tilde{u}_n are solutions of the homogeneous wave equation, i.e., of Eq. (3) with the gain and noise polarization terms omitted, but subject to all the boundary conditions of the optical structure itself. In

practice, these transverse eigenmodes are usually obtained not by solving the differential Eq. (3), but as eigen-solutions of an integral equation¹⁵

$$\tilde{u}_n(\mathbf{s}, z+p) = \int_A \tilde{K}(\mathbf{s}, \mathbf{s}_0, z) \tilde{u}_n(\mathbf{s}_0, z) d\mathbf{s}_0 = \tilde{\gamma}_n \tilde{u}_n(\mathbf{s}, z), \quad (6)$$

where $\tilde{K}(\mathbf{s}, \mathbf{s}_0)$ is a propagation kernel, generally similar to Huygen's integral, but including any finite mirrors, apertures, and intracavity optics in the optical structure. This integral then describes propagation through one period of the empty optical system, starting at any arbitrary reference plane z and propagating to the corresponding reference plane at $z+p$ one period p (or one cavity round trip) later. The differential $d\mathbf{s}_0 \equiv dx_0 dy_0$ is integrated over the full cross section A of the resonator or lens guide at the selected reference plane z .

The eigenvalue $\tilde{\gamma}_n$ gives the complex amplitude reduction and phase shift for the n th order eigenmode after propagation through one period in an optical lens guide or one round trip in an optical resonator. The fractional power loss for the n th mode in one period due to diffraction losses is $1 - |\tilde{\gamma}_n|^2$, while the phase angle of $\tilde{\gamma}_n$ gives the added phase variation over and above the basic $\exp(-j\beta p)$ propagation factor for the same mode.

The integral Eq. (6) directly provides the transverse eigenmodes $\tilde{u}_n(\mathbf{s}, z)$ only at the one arbitrarily chosen reference plane z . These modes of course exist at all other planes z and their (generally slow) variation with z can be determined either by shifting the kernel to a different reference plane z or, more easily, by forward propagation of the modes from the initial reference plane z . The eigenvalues $\tilde{\gamma}_n$ and all the results to be obtained below are entirely independent of the choice of z .

C. Biorthogonality properties

The crucial factor in this analysis is that the propagation integral (6) is in general not a Hermitian operator for any of the usual open-sided structures used as optical resonators or lens guides. One might think that this integral operator should be Hermitian, since it is used to find a solution to the fully Hermitian wave equation. The boundary conditions on the wave equation for open-sided resonators are not Hermitian, however, and this shows up in the integral equation as a non-Hermitian form for the integral operator.

Because the operator is not Hermitian, neither the completeness nor even the existence of a set of eigensolutions \tilde{u}_n to the integral Eq. (6) can be rigorously guaranteed. We must take the existence of such eigenmodes and their usefulness as a basis set, therefore, as matters of empirical (or numerical) observation rather than matters of rigorous mathematical proof. Because of the non-Hermitian character of the integral operator, the transverse modes in these open-sided optical systems are also in general not "normal modes," i.e., they are not power orthogonal or self-adjoint to each other, with the result that

$$\int_A \tilde{u}_n(\mathbf{s}, z) \tilde{u}_m^*(\mathbf{s}, z) d\mathbf{s} \neq \delta_{nm}. \quad (7)$$

Within one period between hard-edged apertures of finite

mirrors, however, we can at least scale the amplitude of each individual mode so that they are individually power normalized, in the sense that

$$\bar{A}_{nn} \equiv \int_A \bar{u}_n(\mathbf{s}, z) \bar{u}_n^*(\mathbf{s}, z) d\mathbf{s} = 1. \quad (8)$$

Of course, as any such mode passes through the finite aperture between successive periods of the structure (or bounces off the output mirror in the cavity case), it loses energy to diffraction losses, so that the power in the mode is reduced by $|\bar{\gamma}_n|^2$ in the next period.

Rather than being orthogonal to each other, the eigenmodes $\bar{u}_n(\mathbf{s}, z)$ are instead biorthogonal to a set of transposed or adjoint eigenfunctions $\bar{\phi}_n(\mathbf{s}, z)$ in the fashion

$$\int_A \bar{u}_n(\mathbf{s}, z) \bar{\phi}_m(\mathbf{s}, z) d\mathbf{s} = \delta_{nm}, \quad (9)$$

where the adjoint eigenmodes $\bar{\phi}_n(\mathbf{s}, z)$ are solutions of the transposed eigenequation

$$\bar{\phi}_n(\mathbf{s}, z) = \int_A \bar{K}^T(\mathbf{s}, \mathbf{s}_0, z) \bar{\phi}_n(\mathbf{s}_0, z) d\mathbf{s}_0 = \bar{\gamma}_n \bar{\phi}_n(\mathbf{s}, z). \quad (10)$$

Here the integral kernel $\bar{K}^T(\mathbf{s}, \mathbf{s}_0, z)$ is the transpose (in \mathbf{s} and \mathbf{s}_0) of the kernel in Eq. (6). The adjoint eigenfunctions $\bar{\phi}_n(\mathbf{s}, z)$ are thus physically different—distinctly different in some cases—from the original eigenfunctions $\bar{u}_n(\mathbf{s}, z)$, although the eigenvalues $\bar{\gamma}_n$ are identical for corresponding modes in the two sets.

In physical terms, it can be shown that if \bar{K} represents the propagation kernel for forward propagation through one period or segment of an optical waveguide, from one arbitrary reference plane to the same plane one period later, then the transposed kernel \bar{K}^T corresponds to propagation in the opposite or reverse direction along the same lens guide, between the same two reference planes. Figure 2 illustrates this difference for a strongly unstable lens guide. The solid lines in this figure indicate the ap-

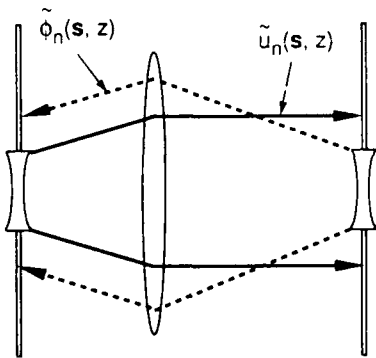


FIG. 2. In an unstable optical lens guide with a large magnification, such as the confocal example shown here, the forward eigenmodes $\bar{u}_n(\mathbf{s}, z)$ will be confined more or less within the volume bounded by the solid lines, while the adjoint eigenmodes $\bar{\phi}_n(\mathbf{s}, z)$ in this example will lie generally within the volume bounded by the dashed lines. Both sets of eigenmodes will, however, have complicated beam profiles with complex Fresnel ripples and some spillover outside the regions indicated. Note that the ordinary and adjoint eigenmodes will both have the same geometric magnification and the same complex eigenvalues $\bar{\gamma}_n$ per period.

proximate outer boundaries for the forward set of propagation eigenmodes $\bar{u}_n(\mathbf{s}, z)$, and the dashed lines the approximate outer boundaries for the reverse set of propagation eigenmodes $\bar{\phi}_n(\mathbf{s}, z)$.

We should note that the adjoint modes $\bar{\phi}_n$ are equally non-normal, i.e.,

$$\int_A \bar{\phi}_n(\mathbf{s}, z) \bar{\phi}_m^*(\mathbf{s}, z) d\mathbf{s} \neq \delta_{nm}. \quad (11)$$

More importantly, these modes cannot be individually power normalized to unity in the same fashion as the \bar{u}_n modes, for suppose we write

$$\bar{\phi}_n(\mathbf{s}, z) \equiv \bar{u}_n^*(\mathbf{s}, z) + \Delta \bar{u}_n^*(\mathbf{s}, z) \quad (12)$$

and also the complex conjugate of this, so that $\Delta \bar{u}_n(\mathbf{s}, z)$ is the difference between a given eigenmode $\bar{u}_n(\mathbf{s}, z)$ and the corresponding $\bar{\phi}_n^*(\mathbf{s}, z)$ having the same eigenmode $\bar{\gamma}_n$. The biorthogonality of \bar{u}_n and $\bar{\phi}_n$ and the power normalization of \bar{u}_n and \bar{u}_n^* then lead to

$$\int_A \bar{u}_n \Delta \bar{u}_n^* d\mathbf{s} = \int_A \bar{u}_n^* \Delta \bar{u}_n d\mathbf{s} = 0, \quad (13)$$

and so the power normalization of the adjoint eigenmodes becomes

$$\int_A \bar{\phi}_n \bar{\phi}_n^* d\mathbf{s} = 1 + \int_A \Delta \bar{u}_n \Delta \bar{u}_n^* d\mathbf{s} \geq 1. \quad (14)$$

This integral is necessarily greater than unity unless the difference $\Delta \bar{u}_n(\mathbf{s}, z) \equiv \bar{\phi}_n^*(\mathbf{s}, z) - \bar{u}_n(\mathbf{s}, z)$ between the regular and adjoint eigenmodes is identically zero.

D. Formal solution

If we substitute the mode expansion of Eq. (5) into the reduced wave Eq. (3), multiply both sides by any one adjoint mode function $\bar{\phi}_n(\mathbf{s}, z)$ and integrate over the transverse cross section A , the wave equation separates into individual equations for the complex amplitude of each transverse mode in the form

$$\frac{d\bar{c}_n(z)}{dz} = \alpha \bar{c}_n(z) - j\bar{p}_n(z), \quad (15)$$

where the Langevin noise term $\bar{p}_n(z)$ that drives each n th mode coefficient $\bar{c}_n(z)$ is given by

$$\bar{p}_n(z) = \frac{\omega \eta_0}{2} \int_A \bar{P}_N(\mathbf{s}, z) \bar{\phi}_n(\mathbf{s}, z) d\mathbf{s}. \quad (16)$$

The formal solution to (15), given an input coefficient $\bar{c}_n(0)$ at $z=0$, is

$$\bar{c}_n(z) = \exp(\alpha z) \bar{c}_n(0) - j \int_0^z \exp[\alpha(z-z')] \bar{p}_n(z') dz'. \quad (17)$$

The crucial point here, as will become more evident later, is that the gain factor α for the n th mode coefficient \bar{c}_n involves an overlap integral of \bar{u}_n and $\bar{\phi}_n$, which evaluates to unity, while the Langevin noise term involves $\bar{\phi}_n$ and $\bar{\phi}_n^*$, whose overlap integral (14) is always greater than or equal to unity. This alters the fundamental relationship between gain and noise emission per unit length for strongly non-Hermitian structures where the transposed eigenfunctions $\bar{\phi}_n(\mathbf{s}, z)$ are distinctly different from the original eigenfunctions $\bar{u}_n^*(\mathbf{s}, z)$.

E. Noise correlations and excess noise factors

The total power $I(z)$ flowing down the lens guide in the $+z$ direction at any plane z can then be written as

$$I(z) = \frac{1}{2\eta_0} \int_A \tilde{E}(s, z) \tilde{E}^*(s, z) ds \\ = \frac{1}{2\eta_0} \sum_n \tilde{c}_n(z) \tilde{c}_n^*(z) + \frac{1}{2\eta_0} \sum_{\substack{n, m \\ n \neq m}} \tilde{A}_{nm} \tilde{c}_n(z) \tilde{c}_m^*(z). \quad (18)$$

The first sum in the second line is the usual sum over the "powers per individual mode," assuming the $\tilde{u}_n(s, z)$ functions to be normalized to unity. The second sum (taken over all values of n and m except $n = m$) expresses the "cross powers" between modes. The constants \tilde{A}_{nm} represent the overlap integrals between different transverse eigenmodes as given by

$$\tilde{A}_{nm} \equiv \int_A \tilde{u}_n(s, z) \tilde{u}_m^*(s, z) ds. \quad (19)$$

These cross-mode integrals can be shown to be independent of axial position z within one period of the lens guide, and the diagonal elements have values $\tilde{A}_{nn} \equiv 1$ for the normalization we have selected.

From Eq. (15), the modes are driven by the Langevin noise polarizations $\tilde{p}_n(z)$ as well as, possibly, by coherent input signals. The coefficients $\tilde{c}_n(z)$ are thus in the general case random variables, so that average powers must be obtained by calculating expectation values of the form $\langle \tilde{c}_n \tilde{c}_n^* \rangle$ or $\langle \tilde{c}_n \tilde{c}_m^* \rangle$. We thus write for the total power at plane z

$$I(z) = \frac{1}{2\eta_0} \sum_n \langle \tilde{c}_n(z) \tilde{c}_n^*(z) \rangle \\ + \frac{1}{2\eta_0} \sum_{\substack{n, m \\ n \neq m}} \langle \tilde{c}_n(z) \tilde{c}_m^*(z) \rangle \tilde{A}_{nm}. \quad (20)$$

But from Eq. (17) the expectation values with the Langevin noise terms included will be given in general by

$$\langle \tilde{c}_n(z) \tilde{c}_m^*(z) \rangle = e^{2\alpha z} \langle \tilde{c}_n(0) \tilde{c}_m^*(0) \rangle \\ + e^{2\alpha z} \int_0^z dz' \int_0^z dz'' e^{-\alpha(z' + z'')} \\ \times \langle \tilde{p}_n(z') \tilde{p}_m^*(z'') \rangle, \quad (21)$$

including the case with $n = m$. In writing this we make the (very reasonable) physical assumption that in a one-way amplifier the spontaneous emission $\tilde{p}_n(z)$ at any plane $z > 0$ will be entirely uncorrelated with the input signal, or noise, contained in the input wave $\tilde{c}_n(0)$ at $z = 0$, i.e., that $\langle \tilde{c}_n(0) \tilde{p}_m^*(z) \rangle \equiv 0$ for $z \geq 0$.

The noise term inside the integral in Eq. (21) can then be written, using Eq. (16), in the form

$$\langle \tilde{p}_n(z') \tilde{p}_m^*(z'') \rangle = \left[\frac{\omega \eta_0}{2} \right]^2 \\ \times \int_A ds' \int_A ds'' \langle \tilde{P}_N(\mathbf{r}') \tilde{P}_N^*(\mathbf{r}'') \rangle \\ \times \tilde{\phi}_n(\mathbf{s}', z') \tilde{\phi}_m^*(\mathbf{s}'', z''). \quad (22)$$

But, since the original noise polarization $\tilde{P}_N(\mathbf{r})$ is δ -function correlated in all spatial coordinates, as given in Eq. (4), we can do the transverse integrations immediately and reduce this to

$$\langle \tilde{p}_n(z') \tilde{p}_m^*(z'') \rangle = 4\rho_I \hbar \omega \eta_0 \alpha B \tilde{B}_{nm} \delta(z' - z''), \quad (23)$$

where we define the \tilde{B}_{nm} to be the (non- z -varying) overlap integrals

$$\tilde{B}_{nm} \equiv \int_A \tilde{\phi}_n(s, z) \tilde{\phi}_m^*(s, z) ds. \quad (24)$$

So long as the adjoint eigenfunctions $\tilde{\phi}_n(s, z)$, like the original eigenfunctions $\tilde{u}_n(s, z)$, are not power orthogonal, the off-diagonal elements \tilde{B}_{nm} for $n \neq m$ will have finite values, and there will be correlation between the Langevin noise terms $\tilde{p}_n(z)$ and $\tilde{p}_m(z)$ driving different transverse eigenmodes in the lens guide.

For the on-diagonal elements, in addition we will adopt the notation

$$K_n \equiv \tilde{B}_{nn} = \int_A \tilde{\phi}_n(s, z) \tilde{\phi}_n^*(s, z) ds \quad (25)$$

and refer to K_n henceforth as the Petermann excess-noise factor for the n th mode. We have already shown that $K_n \geq 1$ in all cases. The value of K_n approaches unity only for near-ideal systems with nearly power-orthogonal modes and near-zero diffraction losses.

F. Noise generation within one period

Putting the Langevin noise sources into Eq. (21) and completing the integration then leads to

$$\langle \tilde{c}_n(z) \tilde{c}_m^*(z) \rangle = e^{2\alpha z} \langle \tilde{c}_n(0) \tilde{c}_m^*(0) \rangle \\ + (e^{2\alpha z} - 1) 2\eta_0 \tilde{B}_{nm} \rho_I \hbar \omega B. \quad (26)$$

Putting this into Eq. (20) gives for the power growth over distance p within one period of the lens guide

$$I(p) = G \left[\sum_n \frac{1}{2\eta_0} \langle \tilde{c}_n(0) \tilde{c}_n^*(0) \rangle \right. \\ \left. + \sum_{\substack{n, m \\ n \neq m}} \frac{1}{2\eta_0} \langle \tilde{c}_n(0) \tilde{c}_m^*(0) \rangle \tilde{A}_{nm} \right] \\ + (G - 1) \rho_I \hbar \omega B \left[\sum_n K_n + \sum_{\substack{n, m \\ n \neq m}} \tilde{A}_{nm} \tilde{B}_{nm} \right], \quad (27)$$

where $G \equiv \exp(2\alpha p)$ is the net power gain due to the atoms within one period from plane z to plane $z + p$. The first two terms obviously give simply the input power $I(z = 0)$ multiplied by the power gain $G \equiv \exp(2\alpha p)$. The

second pair of terms gives the spontaneous emission or noise power generated in and amplified by the laser medium within one period.

Most of the essential results of this analysis are contained in this expression. In particular, the usual equivalent input noise power term $\rho_I \hbar \omega B$ [interpreted as one photon per resolution time, per mode, times the incomplete inversion factor $\rho_I \equiv N_2 / (N_2 - N_1)$] appears in the second term of Eq. (27) as usual, but here it is multiplied by the excess-noise factors K_n for each mode, as well as the dimensionless noise correlation factors \tilde{B}_{nm} . We will introduce one further analytical result before discussing this further.

G. Effects of apertures

The results given thus far apply within one period of the periodic lens guide, in the "free-space" region between two finite apertures or mirrors. At any such apertures the fields are truncated or absorbed and thus suffer diffraction losses. It is simplest, and costs little in generality, to think of the lens guide as having just one such aperture or coupling point per period, with the coordinate system chosen so that $z=0+$ comes just after one such aperture, and $z=p-$ comes just before the next succeeding aperture.

When the wave passes through any such aperture, each

of its eigenmodes is reduced in amplitude by the corresponding eigenvalue amplitude $|\tilde{\gamma}_n|$. We can think of these diffraction losses at the apertures either as reducing the normalized amplitudes \tilde{u}_n of the eigenmodes by the complex eigenvalue $\tilde{\gamma}_n$, and hence the overlap integrals \tilde{A}_{nm} by the amount $\tilde{\gamma}_n \tilde{\gamma}_m^*$ at each successive aperture; or alternatively as reducing the expansion coefficients \tilde{c}_n by the same amount at each successive aperture. We choose the latter approach, so that the normalizations of \tilde{u}_n given earlier will be preserved everywhere. (Note that if we elected to change the normalization of the \tilde{u}_n 's by the ratio $\tilde{\gamma}_n$ on going through each aperture moving in the $+z$ direction, we would correspondingly have to change the ratios of the $\tilde{\phi}_n$'s by the inverse ratio $1/\tilde{\gamma}_n$.)

When the aperture transmission is taken into account, the net gain and net noise generation in one period, going from an input plane at $z=0$ just after one aperture to an output plane at $z=p+$ just after the succeeding aperture, must be computed by extending Eq. (17) with the factor

$$\tilde{c}_n(p+) = \tilde{\gamma}_n \tilde{c}_n(p-) . \quad (28)$$

Note that by definition of the eigenmodes there is no mode conversion at the aperture; the amplitude of each expansion coefficient is merely reduced by its eigenvalue $\tilde{\gamma}_n$. The signal and noise power at the output from one complete period is then written as

$$I(p+) = G \left[\sum_n \frac{1}{2\eta_0} |\tilde{\gamma}_n|^2 \langle \tilde{c}_n(0) \tilde{c}_n^*(0) \rangle + \sum_{\substack{n,m \\ n \neq m}} \frac{1}{2\eta_0} \tilde{\gamma}_n \tilde{\gamma}_m^* \langle \tilde{c}_n(0) \tilde{c}_m^*(0) \rangle \tilde{A}_{nm} \right] + (G-1) \rho_I \hbar \omega B \left[\sum_n |\tilde{\gamma}_n|^2 K_n + \sum_{\substack{n,m \\ n \neq m}} \tilde{\gamma}_n \tilde{\gamma}_m^* \tilde{A}_{nm} \tilde{B}_{nm} \right] , \quad (29)$$

in analogy to Eq. (27).

H. Initial wave excitation factor

Suppose we now think of launching an initial wave or mixture of transverse eigenmodes with a transverse field distribution $\tilde{G}_{in}(\mathbf{s}, t) = \text{Re} \tilde{E}_{in}(\mathbf{s}) e^{j\omega t}$ at an input plane $z=0$ at the input end of the periodic lens guide, so that

$$\tilde{E}_{in}(\mathbf{s}) = \sum_n \tilde{c}_n(0) \tilde{u}_n(\mathbf{s}, 0) . \quad (30)$$

Lacking a completeness proof for the eigenmodes \tilde{u}_n we cannot claim that any arbitrary distribution $\tilde{E}_{in}(\mathbf{s})$ can be expanded in this fashion; but we assume that any useful input wave will be given by such an expansion.

Suppose the input field distribution $\tilde{E}_{in}(\mathbf{s})$ carries unity power, and we wish to achieve the maximum possible initial amplitude $\tilde{c}_q(0)$ for one particular transverse eigenmode with $n=q$. To accomplish this in this particular case, we should not "spatially mode match" the injected signal $\tilde{E}_{in}(\mathbf{s})$ into the desired mode $\tilde{u}_q(\mathbf{s}, 0)$ as is often done in more conventional situations, i.e., we do not want the usual condition that $\tilde{E}_{in}(\mathbf{s}) = \text{const} \times \tilde{u}_q(\mathbf{s}, 0)$. Rather, we should match the input field into the complex conjugate of the *adjoint* mode corresponding to the

desired eigenmode, i.e., the condition for maximum excitation of mode \tilde{u}_q is

$$\tilde{E}_{in}(\mathbf{s}) = (2\eta_0 / \tilde{B}_{qq})^{1/2} \tilde{\phi}_q(\mathbf{s}, 0) , \quad (31)$$

where $\tilde{B}_{qq} \equiv K_q$ is the overlap integral for the ϕ_q function defined in Eqs. (24) and (25).

The initial mode-expansion coefficients $\tilde{c}_n(0)$ will then be, from Eqs. (24) and (25) and the biorthogonality relation,

$$\begin{aligned} \tilde{c}_n(0) &= \int_A \tilde{E}_{in}(\mathbf{s}) \tilde{\phi}_n(\mathbf{s}, 0) d\mathbf{s} \\ &= \begin{cases} (2\eta_0 / \tilde{B}_{qq})^{1/2} \tilde{B}_{nq} , & n \neq q \\ (2\eta_0 \tilde{B}_{qq})^{1/2} \equiv (2\eta_0 K_q)^{1/2} , & n = q . \end{cases} \end{aligned} \quad (32)$$

The total power $I(0+)$ in the system just after the input plane, as calculated using Eq. (18), will necessarily still be unity. The total power in the selected mode $n=q$ by itself will, however, be given by

$$I(0+)_{q \text{ th mode}} = \frac{1}{2\eta_0} |\tilde{c}_q(0)|^2 = K_q . \quad (33)$$

Since $K_q > 1$ in general, there will be *more initial power per mode* put into the selected eigenmode than there is in

the injected signal to start with. This enhanced excitation of the coefficient \tilde{c}_q will necessarily be accompanied by finite excitation of other coefficients \tilde{c}_n , $n \neq q$, and hence there will be still further excess power excitation into all the other eigenmodes.

Conservation of energy is maintained in this situation by the cross-power terms that appear in the second summation in Eq. (18). At least some of these terms will necessarily be negative, so that the total power remains at unity. The excess initial wave excitation of the selected mode $n = q$ is nonetheless real and meaningful, and exactly equal (in power) to the excess noise factor K_q for the same mode.

III. DISCUSSION AND CONCLUSIONS

A. Laser amplifier noise figure

Let us first show that, despite the excess spontaneous emission factor, one can still recover the usual quantum-limited noise figure characteristic of any laser amplifier or, in fact, any other kind of linear phase-preserving signal amplifier. In any given optical resonator or lens-guide structure one particular transverse eigenmode, conventionally designated the $n = 0$ mode, will have smaller diffraction losses and hence a larger eigenvalue $\tilde{\gamma}_0$ than all other eigenmodes. Suppose we arrange to inject maximum intensity into this lowest-loss mode at $z = 0$ using an input wave with total (external) power I_{in} , as described just above, and allow the resulting mode mixture to propagate through multiple sections. We assume the system has net power gain, i.e., $G|\tilde{\gamma}_n|^2 > 1$ at least for the lowest-order mode and possibly also for higher-order modes.

Then, because the lowest-order mode has higher net gain per period than any other eigenmode, after a sufficiently large number of periods N all the higher-order eigenmodes will have small amplitudes relative to the $n = 0$ eigenmode, and the field in the lens guide will become predominantly the $n = 0$ eigenmode. The output power in this case, after N periods, will be given by the cascaded expression

$$\begin{aligned} I(N_p) &= (G|\tilde{\gamma}_0|^2)^N K_0 I_{in} \\ &+ \sum_{k=0}^{N-1} (G|\tilde{\gamma}_0|^2)^k (G-1)|\tilde{\gamma}_0|^2 K_0 \rho_I \hbar \omega B \\ &= G_N K_0 I_{in} + (G_N - 1) \frac{(G-1)|\tilde{\gamma}_0|^2}{G|\tilde{\gamma}_0|^2 - 1} K_0 \rho_I \hbar \omega B, \end{aligned} \quad (34)$$

where $G_N \equiv (G|\tilde{\gamma}_0|^2)^N$ is the overall power gain for the $n = 0$ eigenmode through N complete periods in cascade. The second line of this formula is evidently a generalization, valid for N periodic sections in cascade, of the usual noise figure expression for a single-mode laser amplifier of total gain G_T , namely,

$$I_{out} = G_T I_{in} + (G_T - 1) \rho_I \hbar \omega B. \quad (35)$$

This result displays, however, some significant generalizations.

First, the ratio $(G-1)|\tilde{\gamma}_0|^2 / (G|\tilde{\gamma}_0|^2 - 1)$ in the noise term on the second line simply gives the slightly more complicated (but still standard) noise form that results if the total gain $G_T = G_N \equiv (G|\tilde{\gamma}_0|^2)^N$ of an amplifier results from N noisy amplifier sections of gain G cascaded with noiseless attenuators of transmission $|\tilde{\gamma}_0|^2$ between sections. Second, and more important, the effective input noise power $K_0 \rho_I \hbar \omega B$ appearing with the $G_N - 1$ gain term is the usual ρ_I photons per mode multiplied by the (potentially large) excess-noise factor K_0 for the $n = 0$ mode. But, finally, the input signal power I_{in} from an external signal source is (if properly adjoint coupled) also multiplied by the same factor K_0 , as a result of the "excess initial mode excitation factor." Exactly the usual laser-amplifier noise figure is thus recovered, despite the excess spontaneous emission per mode.

As a practical matter, it is not clear that anyone would want to build a cascaded laser amplifier using an unstable lens-guide structure having large magnification and hence large diffraction losses per period. The above analysis applies equally well, however, to the signal buildup with time in injection-seeded laser oscillators or regenerative amplifiers, and these devices do in practice employ unstable resonator structures. This analysis demonstrates that for an injection-seeded, pulsed oscillator one should first try to make the single-pass gain G reasonably large compared to the mode eigenvalue $|\tilde{\gamma}_0|^2$ in order to optimize noise performance.

But even if this is done, there will be a (large) excess-noise factor K_0 in the effective input noise or initial noise in such structures. In effect, there will be K_0 initial noise photons per mode, rather than just one effective photon, in an unstable injection-seeded laser oscillator. This excess-noise factor per mode can, however, be exactly balanced by the excess initial mode excitation factor K_0 , if adjoint coupling is employed. With adjoint coupling one really can inject more initial energy into one selected cavity eigenmode at $t = 0$ than is available in the external signal one uses as the injection signal.

B. Thermal noise in lossy systems

We consider next the opposite limit of lossy or absorbing systems, and discuss how the excess spontaneous emission should not lead to any violation of usual thermal-equilibrium considerations.

All the results in this paper will apply equally well to a periodic lens guide filled with lossy or absorbing atoms, rather than amplifying atoms, if one simply reverses the signs of σ , α , and the population inversion $N_2 - N_1$ in all the formulas. It may be more convenient in the absorbing case to rephrase the noise power per transverse mode $\rho_I \hbar \omega$ in all the expressions as a thermal noise power given by

$$\rho_I \hbar \omega = \frac{N_2}{N_1 - N_2} \hbar \omega = \frac{\hbar \omega}{e^{\hbar \omega / k T_a} - 1} \equiv k T_{eq}, \quad (36)$$

where T_{eq} is an equivalent temperature which becomes identical to the Boltzmann temperature $T_a \equiv k^{-1} \ln(N_1/N_2)$ in the limit $\hbar \omega \ll k T_a$. We can then

use kT_{eq} as a shorthand for the more general quantum result for thermal excitation of a single mode.

In an absorbing periodic lens guide in thermal equilibrium, not only the lowest-loss eigenmode but in general some substantial number of higher-order, higher-loss eigenmodes will be thermally excited to significant amplitudes by spontaneous emission from the absorbing atoms. To calculate the total steady-state noise power in the lossy case, therefore, all modes and mode cross correlations must be retained. The mode coefficient \tilde{c}_n at the output of, say, the $(k+1)$ -th period of the lens guide, just after the end aperture, can be written, using a combination of Eqs. (15) and (23), as

$$\tilde{c}_n(p+;k+1) = g\tilde{\gamma}_n\tilde{c}_n(p+;k) - jg\tilde{\gamma}_n\tilde{\Lambda}_n^{(k)}, \quad (37)$$

where $g \equiv e^{-\alpha p}$ is the amplitude attenuation due to the atoms in one period. The quantity $\tilde{\Lambda}_n^{(k)}$ given by

$$\tilde{\Lambda}_n^{(k)} \equiv \int_{kp}^{(k+1)p} e^{\alpha z} \tilde{p}_n(z) dz \quad (38)$$

is a Langevin equivalent noise source at the input to the k th period of the lens guide produced by the spontaneous emission within that period. Since the spontaneous emission within different periods of the lens guide will be uncorrelated we can find, using earlier formulas, that

$$\langle \tilde{\Lambda}_n^{(k)} \tilde{\Lambda}_m^{(l)} \rangle = (e^{2\alpha p} - 1) 2\eta_0 \tilde{B}_{nm} k T_{eq} B \delta_{kl}. \quad (39)$$

The mode coefficient \tilde{c}_n after N periods will thus be given by

$$\tilde{c}_n(p+;N) = (g\tilde{\gamma}_n)^N \tilde{c}_n(p+;0) - j \sum_{k=1}^N (g\tilde{\gamma}_n)^k \tilde{\Lambda}_n^{(N-k)} \quad (40)$$

and from Eqs. (26) and (39) we can obtain, after N periods,

$$\begin{aligned} \langle \tilde{c}_n(p+;N) \tilde{c}_m^*(p+;N) \rangle &= \langle (G\tilde{\gamma}_n\tilde{\gamma}_m^*)^N \rangle \langle \tilde{c}_n(p+;0) \tilde{c}_m^*(p+;0) \rangle \\ &+ [1 - \langle (G\tilde{\gamma}_n\tilde{\gamma}_m^*)^N \rangle] \frac{(1-G)\tilde{\gamma}_n\tilde{\gamma}_m^*}{1-G\tilde{\gamma}_n\tilde{\gamma}_m^*} 2\eta_0 \tilde{B}_{nm} k T_{eq} B. \end{aligned} \quad (41)$$

The factor $G \equiv g^2 = e^{-2\alpha p}$ due to the atomic absorption in one period is of course less than unity. After a large enough number of periods so that $(G\tilde{\gamma}_n\tilde{\gamma}_m^*)^N \approx 0$, therefore, Eq. (41) will approach the stationary thermal-equilibrium limit

$$\langle \tilde{c}_n(p+) \tilde{c}_m^*(p+) \rangle = \frac{(1-G)\tilde{\gamma}_n\tilde{\gamma}_m^*}{1-G\tilde{\gamma}_n\tilde{\gamma}_m^*} 2\eta_0 \tilde{B}_{nm} k T_{eq} B. \quad (42)$$

One can then substitute this into (20) to obtain the thermal or blackbody noise power propagating along the lossy line in the steady state.

The periodic structures we are considering consist of periodic segments of atomic medium having power

transmission G and noise temperature T_{eq} , separated by apertures with power transmission $|\tilde{\gamma}_n|^2$ but zero noise temperature. That is, in our model the apertures between periods account for the diffraction losses out the sides of the finite diameter lens guide, and we assume the energy lost in this fashion is simply radiated out into a "cold infinity" with no thermal noise coming back. Hence the apertures themselves emit no thermal noise. There may also in general be lenses, curved mirrors, and other lossless optical elements imbedded in the lossy atomic medium, but these are assumed to be transparent or lossless and hence also contribute no additional noise.

The bracketed ratio $(1-G)\tilde{\gamma}_n\tilde{\gamma}_m^*/(1-G\tilde{\gamma}_n\tilde{\gamma}_m^*)$ appearing in Eq. (42) can then be identified as a generalization of the usual reduction in effective noisiness or apparent noise temperature for a system which contains two loss mechanisms, only one of which has a finite temperature T_{eq} associated with it (cf. Appendix B). Let us assume for simplicity that the atomic loss per section is large enough so that $2\alpha p \gg 1$ or $G \ll 1$. The stationary thermal noise power emerging through any output aperture after many segments of such a lens guide can then be written as

$$I(p+) = kT_{eq} B \left\{ \sum_n |\tilde{\gamma}_n|^2 K_n + \sum_{\substack{n,m \\ n \neq m}} \tilde{\gamma}_n \tilde{\gamma}_m^* \tilde{A}_{nm} \tilde{B}_{nm} \right\}. \quad (43)$$

This evidently expresses a sum over transverse modes in which individual modes may carry more than the normal thermal-equilibrium power $kT_{eq} B$ per mode, since $|\tilde{\gamma}_n|^2 K_n$ may be greater than 1 for strongly unstable systems. There are, however, also cross-power terms of magnitude $\tilde{\gamma}_n \tilde{\gamma}_m^* \tilde{A}_{nm} \tilde{B}_{nm}$ for $n \neq m$, many of which one presumes must have negative values.

It is of interest to consider the relationship between these results and the usual ideas of thermal equilibrium and blackbody radiation in absorbing systems. Consider, for example, a semi-infinite length of a periodic unstable lens guide extending back to $z = -\infty$ and terminating at a final output aperture at $z = 0$ in a zero-temperature half plane, as shown in Fig. 3(a). One can view the final output aperture of this structure as an finite aperture through which one can look back into a more or less black semi-infinite holraum filled with absorbing atoms at temperature T_a .

Equation (43) then tells how much total thermal noise power should be emerging from this aperture, summed over all the transverse modes of the lens-guide structure (assuming for simplicity that the atomic loss per period is large compared to the diffraction loss). For a conventional power-orthogonal waveguide, such as a loss-filled rectangular waveguide with dimensions large compared to a wavelength, it is simple to confirm that Eq. (43) will give exactly the usual thermal or blackbody radiation to be expected from an aperture A looking into a hohlraum at temperature T_a , or from a blackbody surface of the same area with unit emissivity. That is, for an ordinary power-orthogonal waveguide the K_n 's and $\tilde{\gamma}_n$'s will have magnitude unity; the \tilde{A}_{nm} 's and \tilde{B}_{nm} 's will all be zero for

$n \neq m$; and there will be just $2\pi A/\lambda^2$ modes (counting polarizations) in the first summation. Hence the total noise power emerging from the aperture in bandwidth B will be

$$I = \frac{2\pi A}{\lambda^2} kT_{\text{eq}} B = \frac{2\pi A}{\lambda^2} \frac{hfB}{\exp(hf/kT_a) - 1}, \quad (44)$$

which is the usual blackbody result.

One may then ask whether this will also be true for the nonorthogonal example shown in Fig. 3(a), if the summations of Eq. (43) can be evaluated. Haus and Kawakami² were in fact able to verify this for the analogous gain-guided system shown in Fig. 3(b), which is also a

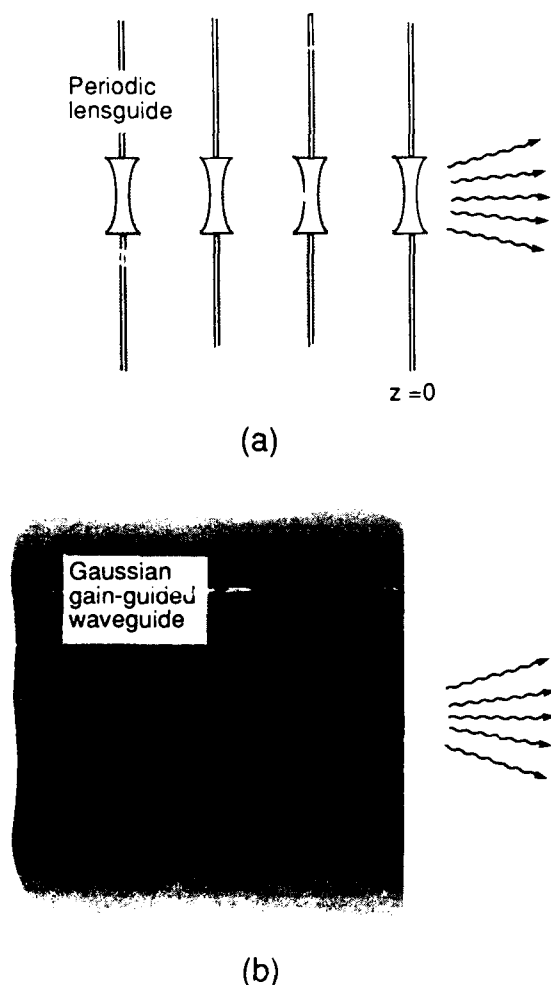


FIG. 3. (a) A semi-infinite length of hard-edged unstable periodic lens guide terminating at an aperture in an infinite plane. The periodic lens guide is filled with absorbing and hence spontaneously emitting atoms at a finite temperature T , but the lenses and dividing walls between sections are at zero temperature. (b) The analogous situation for a semi-infinite length of gain-guided or loss-guided waveguide. The radially varying lossy medium extends out to infinity and also has a uniform finite temperature T .

nonorthogonal or biorthogonal waveguiding system, by appealing to the completeness properties of the infinite set of eigenmodes which one can verify will exist in such a gain-guided structure.

Unfortunately, for the case of more complex hard-edged unstable lens guides such as the one shown in Fig. 3(a), no rigorous proof of the completeness or even the existence of the lens-guide eigenmodes is available; and therefore one is unable at present to confirm that ordinary blackbody emission will be recovered from Eq. (43) for such a structure. From numerical studies one finds empirically that such an unstable lens-guide structure, rather than supporting $\approx 2\pi A/\lambda^2$ transverse modes, seems to possess only a small finite number (say 5–10) of transverse eigenmodes \bar{u}_n with moderate power losses. Moreover, the excess noise factors K_n for these moderate-loss modes are found to be large compared to unity, and to increase very rapidly with increasing mode number n . Beyond this it is not clear whether still higher-order modes exist or not. If they do, they have extremely small eigenvalues, with $|\bar{\gamma}_n| \leq 0.001$, which makes them very difficult to calculate numerically; and they presumably also have very large excess-noise factors. The numerical difficulty in calculating these higher-order terms makes even the purely numerical exploration of Eq. (43) very difficult.

One can be sure from the results of Haus and Kawakami that the cross-mode correlation terms in Eqs. (41) and (43) will be important and that they will remove the apparent violation of thermodynamics produced by the excess noise factors K_n at least for simple gain-guided lens guides. Whether one will recover exactly the usual blackbody results using Eq. (43) for hard-edged unstable lens guides, however, is not clear. Further investigation of the unresolved questions concerning mode properties and noise properties in this situation would be very interesting.

C. Relationship to amplified spontaneous emission

Finally, there is an intriguing question as to what relationship if any there may be between the Petermann excess-noise factor for unstable systems as derived here and a recent discussion of amplified spontaneous emission (ASE) in Cassegrainian amplifiers given by Eimerl.¹⁷ Using purely ray-optic arguments Eimerl notes that in these systems there can be rays of amplified spontaneous emission which make approximately twice as many bounces within the amplifier as do those rays associated with the usual signal beam traveling through the system. Hence there are ASE components which experience the square of the power gain which the usual signal experiences. This excess gain for some of the ASE components can perhaps be interpreted as saying that the net spontaneous emission from the amplifier will be significantly enhanced over its value in conventional geometrically stable or single-pass laser amplifiers, in a fashion which seems very comparable to the Petermann enhancement described in this paper.

ACKNOWLEDGMENTS

The author appreciates the support of the U.S. Air Force Office of Scientific Research for this work, and also the financial support provided by the Alexander von Humboldt Foundation and the Max-Planck-Institute for Quantum Optics in Garching, West Germany, where much of this analysis was done.

APPENDIX A: NOISE POLARIZATION TERM

We give here a brief heuristic derivation of the δ -function-correlated noise polarization term introduced in Eq. (4). The approach followed here is very analogous to that used earlier by Haus¹⁸ and others.

The physical situation envisioned is a collection of two-level atoms with upper- and lower-level population densities N_2 and N_1 . Since the spontaneous emission from each individual atom should be randomly phased and uncorrelated with nearby atoms, we assume this spontaneous emission can be represented by a noise polarization (i.e., a noiselike electric dipole moment per unit volume) whose phasor amplitude, call it $\tilde{P}_N(\mathbf{r})$, will be δ correlated in space in the form

$$\langle \tilde{P}_N(\mathbf{r}) \tilde{P}_N^*(\mathbf{r}') \rangle = C_1 \delta(\mathbf{r} - \mathbf{r}'), \quad (\text{A1})$$

where the correlation volume for the three-dimensional δ function $\delta(\mathbf{r} - \mathbf{r}')$ is of atomic dimensions, and hence much smaller than an optical wavelength.

To determine the constant C_1 we note that the phasor amplitude $\tilde{\mu}_N$ of the oscillating electric dipole moment due to the total spontaneous emission in any small volume element V will be given by

$$\tilde{\mu}_N \equiv \int_V \tilde{P}_N(\mathbf{r}) d\mathbf{r} \quad (\text{A2})$$

where we take the volume element V to be large compared to the correlation volume for $\tilde{P}_N(\mathbf{r})$ but still small compared to an optical wavelength. The average power radiated by a noise dipole $\tilde{\mu}_N e^{j\omega t}$ from the small volume V is then given, from standard em theory, by

$$I_{av} = \frac{\eta_0 \omega^4 \langle \tilde{\mu}_N \tilde{\mu}_N^* \rangle}{12\pi c^2} = \frac{\eta_0 \omega^4 C_1 V}{12\pi c^2}, \quad (\text{A3})$$

where c is the velocity of light in the medium and we have used (A1) in deriving the second equality. But this emission should correspond to the spontaneous emission from the number of upper-level atoms $N_2 V$ in this small volume, assuming each atom radiates at a spontaneous emission rate (or Einstein A coefficient) γ_{rad} . Hence we can also write

$$I_{av} = \frac{\gamma_{rad} N_2 V \hbar \omega}{3}, \quad (\text{A4})$$

where the factor of 3 in the denominator is inserted because the atomic dipoles will in general be randomly polarized and we want the emission only into one sense of polarization, not all polarization directions.

The constant C_1 corresponding to the total spontaneous emission into the full linewidth of a reasonably narrow atomic transition is thus given by

$$C_1(\text{full linewidth}) = \frac{4\pi c^2 \hbar \gamma_{rad} N_2}{\eta_0 \omega^3}. \quad (\text{A5})$$

We are more interested, however, in the noise emission only into a narrow frequency band at the center of the transition. If the transition is Lorentzian with a full width at half maximum (FWHM) linewidth $\Delta\omega_a$, and we consider only a small bandwidth $d\omega \equiv 2\pi B$ at line center, the noise within this bandwidth will be given by

$$C_1(\text{narrow band}) = \frac{d\omega}{\pi \Delta\omega_a / 2} C_1(\text{full linewidth}). \quad (\text{A6})$$

In addition, the midband power gain coefficient 2α for a randomly polarized Lorentzian transition with radiative decay γ_{rad} and FWHM linewidth $\Delta\omega_a$ can be written as

$$2\alpha = \frac{\lambda^2}{2\pi} \frac{\gamma_{rad}}{\Delta\omega_a} (N_2 - N_1). \quad (\text{A7})$$

Combining all of these gives the desired result

$$C_1(\text{bandwidth } B) = \frac{N_2}{N_2 - N_1} \frac{16\hbar\alpha B}{\omega\eta_0}. \quad (\text{A8})$$

Though derived for a Lorentzian transition, this result is in fact very general, for any kind of atomic response producing a net absorption coefficient α at frequency ω .

To verify this we could equally well note that the blackbody-radiation energy density U_{BR} in a narrow bandwidth B in any medium at temperature T can be written as

$$U_{BR} = \frac{\epsilon}{2} |\tilde{E}_{BR}|^2 = \frac{16\pi^2}{\lambda^3} \frac{\hbar B}{\exp(\hbar\omega/kT) - 1}, \quad (\text{A9})$$

where \tilde{E}_{BR} is the phasor amplitude of the blackbody E field in bandwidth B . The energy absorbed from these fields in a volume V of an atomic medium of effective conductivity σ , and hence absorption coefficient $\alpha = \eta_0 \sigma / 2$, will be

$$I_{abs} = \frac{\sigma |\tilde{E}_{BR}|^2 V}{2}. \quad (\text{A10})$$

Equating one-third of this narrow-band power absorption to the power emission of (A3) again gives directly the same result as (A8).

APPENDIX B: TWO-SECTION TRANSMISSION LINE NOISE

Consider a periodic transmission system composed of alternating segments of single-mode transmission line of length l_1 and l_2 having power transmissions $L_1 = \exp(-2\alpha_1 l_1)$ and $L_2 = \exp(-2\alpha_2 l_2)$, and noise temperatures T_1 (finite) and $T_2 = 0$, respectively. Using standard arguments, one can show that the thermal noise power $I(N)$ emerging after N segments of this line, observed at the output end of an l_2 segment, will be

Excess spontaneous emission in non-Hermitian optical systems. II. Laser oscillators

A. E. Siegman

*Max-Planck-Institut für Quantenoptik, D-8046 Garching, West Germany
and E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305*

(Received 15 August 1988)

When the Petermann excess-spontaneous-emission factor described in the accompanying paper [Siegman, preceding paper, *Phys. Rev. A* **39**, 1253 (1989)] is taken into account, we find that the quantum-noise fluctuations or Schawlow-Townes fluctuations in the output spectrum of a laser oscillator will be multiplied by this excess noise factor. The excess spontaneous emission in a laser oscillator cannot be canceled by adjoint-mode-excitation techniques as it can be in a laser amplifier. The resulting noise enhancement can be very sizable (100 to 1000 times or more) in gain-guided laser oscillators, and especially in laser oscillators using geometrically unstable cavities with moderate to high magnifications and Fresnel numbers.

I. INTRODUCTION

We showed in a preceding paper¹ (henceforth referred to as I) that all open-sided optical resonators or lens guides are subject to an excess-spontaneous-emission or excess-quantum-noise factor, sometimes referred to as the Petermann noise factor.² This excess-noise factor arises from the non-self-adjoint or non-power-orthogonal nature of the transverse eigenmodes in these structures. The undesirable effects of this excess-noise emission can be compensated for in laser amplifiers, and in the buildup phase of injection-seeded laser oscillators, by an "excess initial mode excitation factor" which is valid for these same structures, and which has the same numerical value as the excess-spontaneous-emission factor. The purpose of this paper is first to confirm analytically that this excess-noise factor also exists in free-running laser oscillators. More importantly, we show that the compensation mechanism is not relevant in the free-running oscillator case. Hence, all laser oscillators, and especially those using geometrically unstable cavities or gain-guided modes, can be expected to show quantum-noise fluctuations or Schawlow-Townes fluctuations in their output spectra which are enhanced by the excess-noise factor. For gain-guided lasers and unstable-resonator lasers in particular, this excess-noise factor can be quite sizable.

II. ANALYSIS

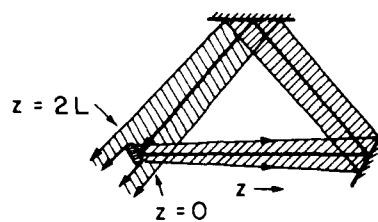
The analytical model in this paper is a resonant laser cavity operating in the oscillating regime, rather than an iterated periodic lens guide or a resonant cavity operating in the transient buildup region as considered in I. We can, however, carry over virtually all the notation and physical concepts of I by thinking of the resonant cavity as consisting of one complete period of a lens guide folded back on itself to form either a ring cavity as in Fig. 1(a), or a standing-wave cavity as in Fig. 1(b). In the latter case the single output mirror or coupling aperture of the structure is taken to be at one end of the cavity, with the z axis for right-traveling waves running from

$z=0$ at this aperture to $z=L$ at the opposite end, and the z axis for left-traveling waves running to the left from $z=L$ to $z=2L \equiv p$ back at the same aperture. This same cavity model has been used in earlier analyses.^{3,4}

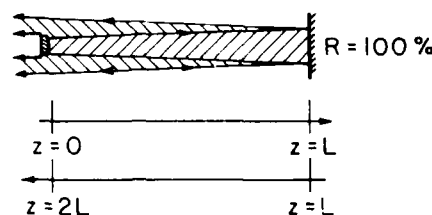
A. Cavity mode analysis

We start with the same wave equation and notations as in Eqs. (1) and (2) of I, except we now assume the phasor amplitudes $\tilde{E}(\mathbf{r}, t)$ and $\tilde{P}_N(\mathbf{r}, t)$ to be functions of position and time rather than position and frequency. The extended paraxial equation for the field in the laser cavity then becomes

$$\left[\nabla_T^2 - 2j\beta_c \left(\frac{\partial}{\partial z} - \alpha + \frac{1}{c} \frac{\partial}{\partial t} \right) \right] \tilde{E}(\mathbf{r}, t) = -\omega_c^2 \mu \tilde{P}_N(\mathbf{r}, t). \quad (1)$$



(a)



(b)

FIG. 1. Analytical models for ring-type and standing-wave laser cavities. Unstable resonators are shown, but any type of single-output-plane resonator could be employed.

We add a subscript to ω_c and β_c to indicate that ω_c is chosen to be a cavity axial mode frequency fixed somewhere near the center of the laser's oscillation spectrum. We then expand $\tilde{E}(\mathbf{r}, t)$ in the actual Fox and Li transverse eigenmodes of the resonator structure in the form

$$\begin{aligned}\tilde{E}(\mathbf{r}, t) &= \sum_n \tilde{a}_n(z, t) \tilde{\gamma}_n \exp[\ln(1/\tilde{\gamma}_n)z/p] \tilde{u}_n(\mathbf{s}, z) \\ &= \sum_n \tilde{a}_n(z, t) \exp[(\alpha_n - j\beta_n)(z-p)] \tilde{u}_n(\mathbf{s}, z),\end{aligned}\quad (2)$$

where the eigenmodes \tilde{u}_n and eigenvalues $\tilde{\gamma}_n$ have all the same properties as in I, and where the traveling-mode amplitude $\tilde{a}_n(z, t)$ is required to obey the boundary condition $\tilde{a}_n(p, t) = \tilde{a}_n(0, t)$. We also introduce the notation

$$\tilde{\gamma}_n \equiv \exp(-\alpha_n p + j\beta_n p) \quad (3)$$

to separate the loss and phase-shift parts $-\alpha_n p$ and $-j\beta_n p$, respectively, of the complex eigenvalue $\tilde{\gamma}_n$.

The advantage of the formulation in Eq. (2) is that with the periodic boundary condition imposed on $\tilde{a}_n(z, t)$, each term in the expansion individually satisfies all the boundary conditions of the laser resonator. That is, for each eigenmode \tilde{u}_n the mode amplitude $\tilde{a}_n(0+, t)$ reflected off the output mirror at time t should be equal to the mode amplitude $\tilde{a}_n(p-, t)$ striking the mirror at the same instant, times the eigenvalue $\tilde{\gamma}_n$ to account for finite reflection from the mirror. The form of (2) guarantees this at all times, and so each individual term in (2) gives the form of the fields in a laser oscillating in that particular transverse mode. We can also note that with this normalization the (one-way) power flow carried by a single transverse mode at plane z is

$$\begin{aligned}I(z, t) &= \frac{1}{2\eta_0} \int_A |\tilde{E}(\mathbf{r}, t)|^2 ds \\ &= \frac{1}{2\eta_0} e^{2\alpha_n(z-p)} |\tilde{a}_n(z, t)|^2\end{aligned}\quad (4)$$

and the power output at the output coupler is then

$$\begin{aligned}I_{\text{out}}(t) &= (1 - |\tilde{\gamma}_n|^2) I(p-, t) \\ &= \frac{1 - |\tilde{\gamma}_n|^2}{2\eta_0} |\tilde{a}_n(p-, t)|^2.\end{aligned}\quad (5)$$

Note, however, that Eqs. (4) and (5) are valid only for a single mode in the cavity, i.e., they do not include the cross-power terms which must be included if more than one mode is present in the cavity.

Suppose the cavity is in fact oscillating in a single axial as well as single transverse mode, so that $\tilde{a}_n(z, t)$ is actually constant with z . The stored energy in the cavity in this one mode (again, ignoring cross-mode terms) is then the integral over the cavity volume

$$W(t) = \frac{\epsilon}{2} \int_V |\tilde{E}(\mathbf{r}, t)|^2 d\mathbf{r} = \frac{\epsilon}{2} \frac{1 - |\tilde{\gamma}_n|^2}{2\alpha_n} |\tilde{a}_n|^2. \quad (6)$$

The "cold cavity" decay rate due to the diffraction losses is then exactly the expected result

$$\gamma_c \equiv \frac{I_{\text{out}}(t)}{W(t)} = \frac{2\alpha_n}{\eta_0 \epsilon} = \frac{2\alpha_n p}{T}, \quad (7)$$

i.e., an exponential power decay coefficient of $2\alpha_n p$ for each cavity round-trip time $T \equiv p/c$.

Substituting (2) into (1), multiplying each side by a specific adjoint eigenmode $\tilde{\phi}_n(\mathbf{s}, z)$ (as previously defined in I), and integrating over the transverse coordinates $d\mathbf{s} \equiv dx dy$ then yields

$$\begin{aligned}\left[\frac{\partial}{\partial z} - (\alpha - \alpha_n + j\beta_n) + \frac{1}{c} \frac{\partial}{\partial t} \right] \tilde{a}_n(z, t) \\ = -j \frac{\omega_c \eta_0}{2} \exp[-(\alpha_n - j\beta_n)(z-p)] \\ \times \int_A \tilde{P}_N(\mathbf{r}, t) \tilde{\phi}_n(\mathbf{s}, z) d\mathbf{s}.\end{aligned}\quad (8)$$

In agreement with the boundary condition on $\tilde{a}_n(z, t)$ given above, the axial dependence of $\tilde{a}_n(z, t)$ can be written as a sum of axial modes of the form $\tilde{c}_{qn} \exp(-jq2\pi z/p)$, where q is an axial mode index. It is then evident from the left-hand side of (8) that each of these axial modes will have a natural resonance frequency ω_{qn} given by

$$\omega_{qn} = q(2\pi c/p) + \beta_n c. \quad (9)$$

The index q thus identifies successive axial modes spaced apart by $\Delta\omega_{ax} = 2\pi c/p$. The $\beta_n c$ term accounts for the additional, normally small, frequency shifts associated with different transverse modes. We thus write $\tilde{a}_n(z, t)$ in the general form

$$\begin{aligned}\tilde{a}_n(z, t) &= \sum_q \tilde{c}_{qn}(t) \exp\{j[q(2\pi c/p) + \beta_n c]t \\ &\quad - jq2\pi z/p\}.\end{aligned}\quad (10)$$

Putting this into (8), multiplying both sides by $\exp(+jq2\pi z/p)$, and integrating over z then gives, finally, for each individual axial modes component, the equation of motion

$$\left[\frac{d}{dt} - (\hat{\alpha} - \hat{\alpha}_n) \right] \tilde{c}_{qn}(t) = -j\tilde{p}_{qn}(t), \quad (11)$$

where we use $\hat{\alpha} \equiv \alpha c$ and $\hat{\alpha}_n \equiv \alpha_n c$ to indicate growth rates in time rather than space. The quantity $\tilde{p}_{qn}(t)$ is the Langevin noise term driving this particular axial and transverse mode, as given by

$$\begin{aligned}\tilde{p}_{qn}(t) &= \frac{\omega_c c \eta_0}{2p} e^{-j\omega_{qn} t} \int_0^p dz e^{jq2\pi z/p} e^{-(\alpha_n - j\beta_n)(z-p)} \\ &\quad \times \int_A \tilde{P}_N(\mathbf{r}, t) \tilde{\phi}_n(\mathbf{s}, z) d\mathbf{s}.\end{aligned}\quad (12)$$

It remains only to determine the properties of this Langevin noise term.

B. Langevin noise terms

We are in most cases interested only in those situations where the linewidth of the atomic transition will be much wider than the oscillation bandwidths or the growth or decay rates of any oscillating mode in the laser cavity. This is the same as saying that the noise polarization $\tilde{P}_N(\mathbf{r}, t)$ due to spontaneous emission will have a very

short correlation time compared to all other time constants of interest in the problem, so that $\bar{P}_N(\mathbf{r}, t)$ can be viewed as being effectively δ -function correlated in time as well as in spatial coordinates. From an analysis similar to that in Appendix A in I, we can find that the spontaneous emission polarization obeys the relationship

$$\langle \bar{P}_N(\mathbf{r}, t) \bar{P}_N^*(\mathbf{r}', t') \rangle = \frac{16\rho_I \hbar \alpha}{\omega \eta_0} \delta(t - t') \delta(\mathbf{r} - \mathbf{r}'), \quad (13)$$

where $\rho_I \equiv N_2 / (N_2 - N_1)$ represents possible incomplete inversion on the transition, as discussed in I. An explicit example of the temporal correlation function for the particular case of a Lorentzian atomic line shape would be

$$\langle \bar{P}_N(\mathbf{r}, t) \bar{P}_N^*(\mathbf{r}', t') \rangle = \frac{4\rho_I \hbar \Delta\omega_a \alpha}{\omega \eta_0} e^{-|t - t'|/T_2} \delta(\mathbf{r} - \mathbf{r}'), \quad (14)$$

where $\Delta\omega_a$ is the full width at half maximum (FWHM) linewidth and $T_2 \equiv 2/\Delta\omega_a$ the dephasing time for the atomic linewidth. In the usual limit of T_2 very much shorter than all other time constants, the exponential function in Eq. (14) converts into the temporal δ function of Eq. (13).

Note that $\bar{P}_N(\mathbf{r}, t)$ in the preceding expressions is to be interpreted as the peak (not rms) phasor amplitude of the broadband noise polarization centered at frequency ω_c . Combining (12) and (13) then gives for the Langevin noise term driving the qn th axial and transverse mode

$$\langle \bar{p}_{qn}(t) \bar{p}_{qn}^*(t') \rangle = \left[\frac{\omega_c \eta_0}{2p} \right]^2 \frac{1 - |\bar{\gamma}_n|^2}{2\alpha |\bar{\gamma}_n|^2} \frac{16\rho_I K_n \hbar \alpha}{\omega \eta_0} \times \delta(t - t'), \quad (15)$$

where $K_n \equiv \int_A \bar{\phi}_n \bar{\phi}_n^* ds$ is again the excess noise factor for the n th transverse mode as introduced in I.

C. Single-mode oscillator analysis

Suppose we have a single-mode laser oscillator which oscillates only in a single preferred qn th mode, with all other axial and transverse modes remaining below threshold. The equation of motion for this mode is then the Langevin equation (11), i.e.,

$$\left[\frac{d}{dt} - (\hat{\alpha} - \hat{\alpha}_n) \right] \bar{c}_{qn}(t) = -j\bar{p}_{qn}(t). \quad (16)$$

We need not consider the solutions to this equation in detail, since Langevin equations of this form have been the subject of many previous analyses.⁵⁻¹¹ We note, however, that if the Langevin noise term $\bar{p}_{qn}(t)$ is δ correlated in time in the form

$$\langle \bar{p}_{qn}(t) \bar{p}_{qn}(t') \rangle = N_0 \delta(t - t'), \quad (17)$$

where the constant N_0 is given by (15), then the amplitude of the qn th mode will grow with time according to the formula

$$\frac{d}{dt} \langle \bar{c}_{qn} \bar{c}_{qn}^* \rangle = 2(\hat{\alpha} - \hat{\alpha}_n) \langle \bar{c}_{qn} \bar{c}_{qn}^* \rangle + N_0. \quad (18)$$

Recalling that the total stored energy in the qn th cavity mode will be given by

$$W_{qn} = \frac{\epsilon}{2} \frac{1 - |\bar{\gamma}_n|^2}{2\alpha_n} \langle \bar{c}_{qn} \bar{c}_{qn}^* \rangle, \quad (19)$$

we can convert (18) into the form

$$\frac{dW_{qn}}{dt} = 2(\hat{\alpha} - \hat{\alpha}_n) W_{qn} + 2\hat{\alpha} K_{qn} \hbar \omega, \quad (20)$$

where K_{qn} is a dimensionless excess-noise factor for the qn th axial and transverse mode given by

$$K_{qn} \equiv \left[\frac{1}{|\bar{\gamma}_n|^2} \left[\frac{1 - |\bar{\gamma}_n|^2}{2\alpha_n p} \right]^2 \right] K_n, \quad (21)$$

with $K_n \equiv \int_A \bar{\phi}_n \bar{\phi}_n^* ds$ being the Petermann excess-noise factor for the n th transverse mode.

III. DISCUSSION AND CONCLUSIONS

A. Laser oscillator noise

The important results of this paper are summarized in Eqs. (15) through (21). As already noted, Eq. (16) is the standard Langevin equation for a single-mode van der Pol oscillator. Discussions of the oscillator noise properties resulting from this equation (with an appropriate model for saturation of the gain coefficient $\hat{\alpha}$ taken into account) can be found in numerous papers in the literature, and need not be repeated here.

The important point brought out here is that the Langevin noise amplitude commonly assumed in earlier laser analyses must actually be multiplied by the noise enhancement factor K_{qn} given in Eq. (21), where this noise enhancement factor in general has a value greater than unity. To emphasize this point, we can express the energy in the qn th oscillating mode as a number of photons $n_{qn} \equiv W_{qn}/\hbar\omega$. Equation (20) can then be rearranged into the form

$$\frac{dn_{qn}}{dt} = \kappa N_2 (n_{qn} + K_{qn}) - \kappa N_1 n_{qn} - \gamma_c n_{qn}, \quad (22)$$

where κ is a coupling term between atoms and the cavity mode and γ_c is the cold cavity decay rate. The single "extra photon" which normally appears in the first term of this equation in the form $\kappa N_2 (n_{qn} + 1)$ is now replaced by $\kappa N_2 (n_{qn} + K_{qn})$, i.e., there appear to be K_{qn} extra photons, where K_{qn} is given by (21). All the contributions to oscillator noise fluctuations and spectral broadening resulting from spontaneous emission in all previous laser noise analyses, must thus be multiplied by this K_{qn} factor.

The noise factor K_{qn} in Eq. (21) has two independent parts. The factor contained in the large square brackets is a standard classical result which describes how the effective Langevin noise source is magnified in a cavity with very large output coupling (note that the eigenvalue $|\bar{\gamma}_n|^2$ appearing in this expression could be replaced by the output mirror reflectivity R in most laser cavities). This factor is entirely independent of the transverse-mode biorthogonality properties, and has been derived as an

additional factor multiplying the usual Schawlow-Townes noise expression in a number of earlier laser analyses.^{7,8} The value of this factor remains very close to unity unless the cavity output coupling becomes very large ($\geq 95\%$ /pass), and hence this factor can be neglected for all except the largest output couplings.

The remaining factor K_n is the Petermann excess-spontaneous-emission factor arising from the transverse-mode orthogonality properties as discussed in I. Since this factor is inherently built into the Langevin noise term which drives the oscillating laser field, it is clear there is no chance of canceling this excess emission with any "excess mode excitation factor" as could be done for the amplifier case treated in I.

B. Noise correlation between different transverse modes

We could write another expression similar to Eq. (15) for the noise correlation between the Langevin terms driving different axial and transverse modes, e.g., terms of the form $\langle \tilde{p}_{qn}(t) \tilde{p}_{rm}'(t') \rangle$, where q and r indicate different axial and n and m different transverse modes. Overlap integrals between different modes of the form $\tilde{B}_{nm} = \int \tilde{\phi}_n \tilde{\phi}_m^* ds$ would then arise as they did in I.

The situation is more complicated in the oscillator case as compared to the amplifier or lensguide case, however. For one thing, we are usually interested in practice in single-mode oscillators which oscillate in just a single preferred axial and transverse mode, with all other modes remaining below threshold. Correlations between the spontaneous emission into different modes are then not of interest. In addition, in a resonant laser cavity such as we are considering here the different transverse modes, as well as the different axial modes, will in general have distinctly different resonance frequencies. In the amplifier case we may be concerned with signals in multiple transverse modes, but all within an arbitrarily narrow bandwidth centered about the same frequency ω . Correlations between the noise voltages emitted into the different transverse modes at the same frequency are then relevant and of interest.

In the oscillator case, where each mode has a different frequency, the spontaneous emission into different modes will to first order be uncorrelated, because there is little or no correlation between different frequencies in the spontaneous emission itself. This absence of correlation shows up mathematically in the resonance frequency term $\exp(-j\omega_{qn}t)$, as well as the axial mode variation term $\exp(jq2\pi z/p)$, in Eq. (12). These terms will not cancel out in evaluating a cross-correlation term $\langle \tilde{p}_{qn} \tilde{p}_{rm}^* \rangle$, as they do in evaluating Eq. (15). The result is that cross correlations between different modes will average out over a time corresponding to the reciprocal frequency difference between them.

C. Noise fluctuations in semiconductor lasers

We might emphasize again that the excess noise factor K_n discussed in this paper represents a real increase in the spontaneous-emission rate into each cavity mode. This increased emission rate if present in a semiconduc-

tor diode laser should give a noise enhancement entirely separate and distinct from the "linewidth enhancement factor" $1 + \alpha_c^2$ for semiconductor lasers discussed by Henry⁹⁻¹¹ and others,¹² where α_c is a factor relating changes in refractive index to changes in laser gain for the semiconductor laser medium.

As we noted in I, Petermann² originally calculated an excess spontaneous emission factor K applicable to semiconductor lasers (or indeed any other kind of lasers) in which the transverse mode profile is determined by gain guiding, or alternatively by loss guiding, inside the laser cavity. Some, but by no means all, semiconductor diode lasers use gain guiding to confine the propagating mode in the directions either perpendicular to the junction plane or in the junction plane; and these lasers will exhibit a Petermann excess-emission factor. Other semiconductor diode laser designs rely instead on index guiding for mode confinement; pure index guiding should not by itself lead to a Petermann excess-noise factor.

We have now shown in I, however, that an essentially similar excess-emission factor K will also occur in any type of laser, semiconductor or otherwise, that makes use of geometrically unstable cavity optics. For simplicity we did not include either gain or index guiding effects in this analysis, although the combined effects of gain guiding and geometrically unstable cavity optics could fairly readily be calculated also. Wide-stripe semiconductor lasers in particular could well be operated using external resonator optics which are geometrically unstable in the plane of the diode junction, together with either gain or index guiding being present perpendicular to the junction plane. Such lasers would then exhibit a total excess-spontaneous-emission factor K given by Eq. (25) of I, where the biorthogonality of the cavity mode functions $\tilde{\phi}_n(s, z)$ would result from a combination of the gain-guiding and unstable-resonator effects.

In any semiconductor laser where the physical conditions leading to Henry's α_c factor are present, spontaneous emission into the cavity modes will cause phase or frequency fluctuations in the laser output signal in two ways: both directly through quadrature spontaneous emission events which shift the instantaneous oscillation phase directly, and also indirectly through in-phase spontaneous emission events which change the oscillation amplitude, hence the instantaneous laser gain, and hence the instantaneous cavity frequency. Since spontaneous emission is the direct cause of both of these effects, one might expect the total quantum linewidth of the laser to be multiplied both by the excess-spontaneous-emission factor K discussed in these papers (whether this comes from gain guiding or unstable optics or both), and again multiplied by the linewidth enhancement factor $1 + \alpha_c^2$ predicted by Henry. The usual Schawlow-Townes result for quantum-limited laser linewidth should thus be multiplied by the product of both of these effects.

D. Experimental relevance

Finding numerical values for the excess emission factor K_n requires detailed knowledge of the eigenmodes $\tilde{u}_n(s, z)$ and the adjoint eigenmodes $\tilde{\phi}_n(s, z)$ of the laser cavity.

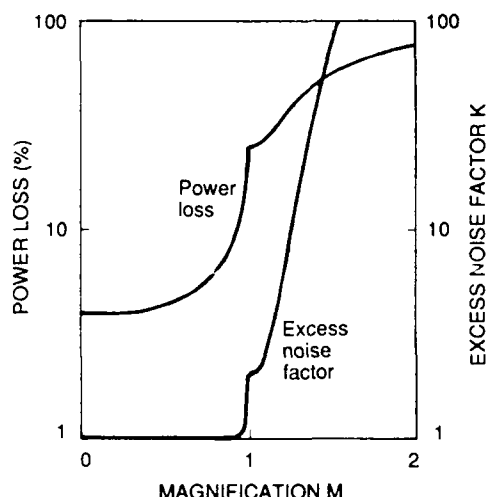


FIG. 2. Round-trip power loss per pass and excess-noise factor for the lowest-order Hermite-Gaussian mode in a standing-wave laser cavity with a Gaussian variable-reflectivity output mirror. Between 0 and 1 the horizontal scale is the half-trace parameter $m = (A + D)/2$, where A and D are the round-trip real ray matrix elements of the cavity. Beyond $M=1$ the horizontal scale is the usual geometric magnification $M \equiv m + (m^2 - 1)^{1/2}$. The variable-reflectivity mirror has aperture spot size given by w_{Ga} . $w_c = 5$, where w_{Ga} is the Gaussian parameter for the variable-reflectivity mirror profile and w_c is the confocal parameter characteristic of the cavity length.

These eigenmodes are notoriously difficult to calculate for conventional hard-edged unstable resonators, and generally require extensive numerical computations. The Hermite-Gaussian eigenmodes that develop in resonators using either stable or unstable cavity optics, combined with Gaussian variable-reflectivity mirrors, can, however, quite easily be found¹³ and the excess-emission factors can readily be determined.¹⁴

Figure 2 shows, for example, both the round trip power loss per pass and the excess-noise factor for the lowest-order Hermite-Gaussian mode in a standing-wave

laser cavity having a single Gaussian variable-reflectivity output mirror, as a function of the degree of magnification or instability of the cavity optics. The region between 0 and 1 on the horizontal axis is the geometrically stable region, with $M=0$ representing a confocal resonator and $M=1$ a plane-parallel resonator. The horizontal scale beyond $M=1$ is the usual unstable-resonator geometric magnification M . The excess-emission factor for this case (and all other similar cases) is near unity in the stable region; rises close to 2 for the planar-resonator case at the stable-unstable boundary; and increases rapidly with geometric magnification in the unstable region. Other preliminary calculations^{14,15} show that much the same thing happens, and very large excess-noise factors are obtained, with conventional hard-edged unstable resonators as well.

The analysis in this paper has of course neglected many higher-order effects also present in real laser oscillators, e.g., finite distribution of the gain medium within the laser cavity, gain saturation, mode-beating effects, spatial hole burning, cross coupling between modes, and so forth. It is believed, however, that the basic conclusions of this paper are of a fundamental nature and will remain valid even if somewhat modified when these other effects are taken into account.

It would be very interesting at this point to carry out experimental tests to measure the excess-noise factor and the excess-quantum-noise fluctuations in a laser oscillator, especially in a system employing an unstable optical resonator. It would be desirable to employ an advantageous (i.e., quieter) gain medium than a gain-guided semiconductor laser in carrying out such tests. Planning toward this objective is in progress.

ACKNOWLEDGMENTS

This work was carried out with support from the U.S. Air Force Office of Scientific Research, and was greatly aided by the hospitality of the Max-Planck-Institut für Quantenoptik in Garching, West Germany, and the support of the Alexander von Humboldt Foundation.

¹A. E. Siegman, preceding paper, Phys. Rev. A **39**, 1253 (1989).

²K. Petermann, IEEE J. Quantum Electron. **QE-15**, 566 (1979).

³A. E. Siegman, Opt. Commun. **31**, 369 (1979).

⁴A. E. Siegman, Appl. Phys. Lett. **36**, 412 (1980).

⁵M. Sargent III, M. O. Scully, and W. E. Lamb, Jr., *Laser Physics* (Addison-Wesley, Reading, MA, 1974).

⁶Jan Perina, *Coherence of Light* (Reidel, Boston, 1985), Sec. 17.2.

⁷K. Ujihara, IEEE J. Quantum Electron. **QE-20**, 814 (1984).

⁸C. H. Henry, J. Lightwave Technol. **LT-4**, 288 (1986).

⁹C. H. Henry, IEEE J. Quantum Electron. **QE-18**, 259 (1982).

¹⁰C. H. Henry, IEEE J. Quantum Electron. **QE-19**, 1391 (1983).

¹¹C. H. Henry, J. Lightwave Technol. **LT-4**, 298 (1986).

¹²M. Osinski and J. Buus, IEEE J. Quantum Electron. **QE-23**, 9 (1987).

¹³A. E. Siegman, *Lasers* (University Science Books, Mill Valley, CA, 1986); Sec. 23.3.

¹⁴A. E. Siegman, P. L. Mussche, and J.-L. Dumont (unpublished).

¹⁵M. A. Lauder and G. H. C. New (private communication).

$$I(N) = (L_1 L_2)^N I(0) + [1 - (L_1 L_2)^N] \frac{(1 - L_1) L_2}{1 - L_1 L_2} k T_1 B$$

$$\approx \frac{(1 - L_1) L_2}{1 - L_1 L_2} k T_1 B \quad \text{if } L^N \ll 1. \quad (\text{B1})$$

This formula is the single-transverse-mode analog of Eqs. (41) and (42). It leads to the three limiting cases

$$I_N \approx \begin{cases} k T_1 B, & \alpha_2 l_2 \rightarrow 0, L_2 \rightarrow 1 \\ L_2 k T_1 B, & \alpha_1 l_1 \gg 1, L_1 \rightarrow 0 \\ \frac{\alpha_1 l_1}{\alpha_1 l_1 + \alpha_1 l_2} k T_1 B, & \alpha_1 l_1, \alpha_2 l_2 \ll 1, L_1, L_2 \rightarrow 1 \end{cases} \quad (\text{B2})$$

and all three of these limits are obviously physically correct.

¹K. Petermann, IEEE J. Quantum Electron. **QE-15**, 566 (1979).

²H. A. Haus and S. Kawakami, IEEE J. Quantum Electron. **QE-21**, 63 (1985).

³Y. Suematsu and K. Furuya, Trans. IECE Jpn. **E60**, 467 (1977).

⁴W. Streifer, D. R. Scifres, and R. D. Burham, Electron. Lett. **17**, 933 (1981).

⁵E. Patzak, Electron. Lett. **18**, 278 (1982).

⁶D. Marcuse, Electron. Lett. **18**, 920 (1982).

⁷A. Yariv and S. Margalit, IEEE J. Quantum Electron. **QE-18**, 1831 (1982).

⁸T. Mamine, J. Appl. Phys. **54**, 2103 (1983).

⁹J. Arnaud, Electron. Lett. **19**, 688 (1983).

¹⁰J. Arnaud, Electron. Lett. **19**, 798 (1983).

¹¹G. P. Agrawal, J. Opt. Soc. Am. Commun. **1**, 406 (1984).

¹²M. Newstein, IEEE J. Quantum Electron. **QE-20**, 1270 (1984).

¹³M. B. El Mashade and J. Arnaud, IEEE J. Quantum Electron. **QE-22**, 505 (1986).

¹⁴C. H. Henry, J. Lightwave Technol. **LT-4**, 288 (1986).

¹⁵See, for example, A. E. Siegman, *Lasers* (University Science Books, Mill Valley, CA, 1986), Chaps. 14–23, especially Sec. 21.7.

¹⁶A. E. Siegman, Opt. Commun. **31**, 369 (1979).

¹⁷David Eimerl, Appl. Opt. **26**, 1594 (1987).

¹⁸H. A. Haus, J. Appl. Phys. **32**, 493 (1961).

(revised 3/5/1989)

Excess Spontaneous Emission in Variable-Reflectivity-Mirror Lasers

Jean-luc Doumont, Paul L. Mussche, and A. E. Siegman

*E. L. Ginzton Laboratory, Stanford University,
Stanford, California 94305-4085*

(Received 1989)

Abstract—We calculate the so-called excess spontaneous emission factors, or excess noise factors, for the lowest- and higher-order transverse modes in geometrically stable and unstable laser cavities having gaussian variable-reflectivity mirrors. These excess emission factors determine both the Schawlow-Townes linewidths or quantum noise fluctuations in laser oscillators and the initial noise levels from which oscillation buildup starts in regenerative laser cavities. The excess emission factor for the lowest-order gaussian mode is found to be near unity for geometrically stable resonators; rises to a value very close to 2 at the stable-unstable boundary (planar-mirror case); and becomes substantially larger than unity for cavities having geometrically unstable optics. Excess emission factors as large as 10 to 100 times and larger are obtained in cavities having geometric magnifications of order 1.5 to 2 and aperture Fresnel numbers of a few times unity, increasing quadratically for higher values of the Fresnel number or the square of the magnification. Higher-order modes have even larger excess noise factors.

I. INTRODUCTION

Spontaneous emission from upper-level atoms is a fundamental and unavoidable source of quantum noise in laser devices. This quantum noise emission is often described as having a magnitude corresponding, if properly expressed, to one excess photon or one quantum noise photon per mode in the laser cavity. Some years ago, however, Petermann calculated an "excess spontaneous emission factor" which predicted an excess spontaneous emission rate in gain-guided semiconductor lasers [1]. Petermann's analysis predicted an excess spontaneous emission rate or excess quantum noise level corresponding to K_p photons per cavity mode, where K_p could be greater than unity in gain-guided semiconductor laser structures. Other workers then made similar predictions. Haus and Kawakami in particular clarified and explained some apparent thermodynamic problems associated with such an excess spontaneous emission rate [2]. For a more extensive set of references, see the list of citations in papers [3] and [4].

More recently, it was realized by one of us that essentially the same sort of excess spontaneous emission would occur in any sort of open-sided optical resonator or lensguide, as a consequence in essence of the non-power-orthogonal nature of the transverse modes in such structures. Two papers analysing this excess spontaneous emission process, and describing the resulting quantum noise fluctuations in laser amplifiers [3] and oscillators [4], have recently been published. One general result from this analysis is that the excess emission factor K_p remains close to unity (i.e., there is little or no excess spontaneous emission) for index-guided lensguides or geometrically stable optical cavities; but K_p can become substantially greater than unity for gain-guided lensguides or for geometrically unstable laser cavities.

The primary objective of this paper is to provide analytical formulas and

numerical values for the excess noise factors, and also for certain other noise correlation factors between different modes, in geometrically stable and unstable lensguides and optical cavities which employ gaussian variable-reflectivity mirrors or "soft" gaussian apertures. We would also like to have similar numerical values for conventional hard-edged unstable resonators. However, the transverse eigenmodes in conventional hard-edged unstable resonators are very complicated, so that extensive numerical calculations are necessary to find these eigenmodes and calculate the resulting excess emission factors. A few such results for hard-edged unstable resonators have recently been reported by Lauder and New [5]; and we will report additional results for such systems in a later publication. The transverse modes of optical resonators which employ gaussian variable-reflectivity mirrors or "soft" gaussian apertures can, on the other hand, be determined analytically, even when the underlying resonator optics are geometrically unstable; and hence the excess emission factors in these cavities can be calculated and expressed in analytic form. Resonators of this form are also of increasing interest for practical laser systems, so that such results may be of practical significance.

In this paper we present in abbreviated form a simple analysis using complex-valued $ABCD$ matrices to calculate the transverse eigenmodes and the excess emission factors for the lowest and higher-order modes in gaussian variable-reflectivity mirror (VRM) systems, both stable and unstable. The basic eigenmode properties of VRM resonators are by now fairly well understood [6]. The eigenmodes in VRM cavities have the form of complex Hermite-gaussian functions with, in general, different complex arguments in the gaussian exponent and the Hermite polynomial functions. We show in this paper that VRM cavities of this sort have excess emission factors significantly greater than unity whenever the internal optics

of the cavity are geometrically unstable. The excess noise factors increase very rapidly with increases in either the geometrical magnification M , the aperture Fresnel number N_{ga} , or the Hermite-gaussian mode indices.

II. ANALYSIS

A. Cavity Mode Analysis

We take as our analytical model a standing-wave type variable-reflectivity mirror (VRM) laser cavity, such as that shown schematically in Fig. 1a. The cavity is assumed to have a gaussian variable-reflectivity mirror at the right-hand end and a 100% reflecting mirror at the left-hand end. Between the end mirrors there can be an arbitrary assortment of lossless paraxial elements, lenses, etc. The one-way transmission through all these paraxial elements, from left to right, can be represented by a real-valued ray matrix or $ABCD$ matrix with elements $A = a$, $B = b$, $C = c$, and $D = d$. Figure 1b shows the lensguide equivalent to this cavity. The $ABCD$ matrix approach to laser resonators is reviewed in laser textbooks such as [6].

The amplitude reflectivity $\rho(r)$ from the gaussian mirror is assumed to have the form

$$\rho(r) = \exp[-r^2/w_{ga}^2] \quad (1)$$

so that w_{ga} is the $1/e$ amplitude width of the equivalent gaussian aperture. The power reflectivity from the mirror is this amplitude reflectivity squared. Reflection from this mirror can be represented in the $ABCD$ formalism by a complex-valued $ABCD$ matrix with elements $A = 1$, $B = 0$, $C = -j\epsilon$, and $D = 1$ where $\epsilon \equiv \lambda/\pi w_{ga}^2$ is $\ll 1$ in practical situations.

To calculate the excess noise factor, one needs to know the resonator transverse eigenmodes $\tilde{u}(x, y)$ or $\tilde{u}(r, \theta)$, and a set of so-called adjoint eigenmodes $\tilde{\phi}(x, y)$ which represent physically the lensguide eigenmodes traveling in the reverse direction. For the analytical model considered here, it is most convenient to evaluate these eigenmodes on the surface of the 100% mirror at the *opposite* end of the cavity from the VRM (see Fig. 1b), since this is a symmetry plane where the transverse and adjoint eigenmodes $\tilde{u}(x, y)$ and $\tilde{\phi}(x, y)$ have the same form (though not the same amplitude). We have represented the one-way propagation from the 100% mirror surface to the gaussian aperture by a purely real ray matrix with elements $A = a$, $B = b$, $C = c$, and $D = d$ going from left to right. The matrix elements going in the reverse direction from the gaussian aperture back to the 100% reflecting mirror are then $A = d$, $B = b$, $C = c$, and $D = a$. Note that the 100% mirror can without loss of generality be taken as a plane mirror, since any curvature of this mirror can be absorbed as an equivalent thin lens in the $abcd$ matrix elements.

The complete complex-valued round-trip ray matrix in the VRM cavity is then given by

$$\begin{aligned} \begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} &= \begin{bmatrix} d & b \\ c & a \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -j\epsilon & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} ad + bc - jabe\epsilon & 2bd - jb^2\epsilon \\ 2ac - ja^2\epsilon & ad + bc - jabe\epsilon \end{bmatrix} \\ &= \begin{bmatrix} \tilde{m} & (\tilde{m} + 1)B/(m + 1) \\ (m + 1)(\tilde{m} - 1)/B & \tilde{m} \end{bmatrix} \end{aligned} \quad (2)$$

where $B \equiv 2bd$ is an effective length for the cavity; $m \equiv ad + bc$ is the half-trace of the real part of the round-trip $ABCD$ matrix in the cavity; $\delta \equiv ab\epsilon = ab\lambda/\pi w_{ga}^2$ is a dimensionless measure of the gaussian aperture contribution; and $\tilde{m} \equiv m - j\delta$ is the complex half-trace value for the complete round trip. From the standard complex-paraxial resonator theory [6], cavities with $-1 < m < 1$ are geometrically

stable, while cavities with $m^2 > 1$ are geometrically unstable, either positive-branch ($m > 1$) or negative-branch ($m < -1$). Note that we use superimposed tildes to indicate that the matrix elements \tilde{A} , \tilde{B} , \tilde{C} and \tilde{D} will in general be complex-valued.

Using the complex-paraxial resonator analysis of [6], one can solve for the gaussian modes of this cavity under ordinary stable conditions, that is, with $m^2 < 1$ and without variable reflectivity on the right-hand mirror (so that $w_{ga} \rightarrow \infty$). The spot size on the right-hand (gaussian) mirror in this case will be given by $w^2 = 2ab\lambda/\pi\sqrt{1-m^2}$. In the confocal resonator limit, which corresponds to $m = 0$, this spot size will have the minimum value $w_{cf}^2 = 2ab\lambda/\pi$. The gaussian mirror parameter δ can thus be given a physical interpretation in terms of the ratio of this minimum or confocal resonator spot size w_{cf} to the gaussian aperture size w_{ga} in the form

$$\delta \equiv \frac{ab\lambda}{\pi w_{ga}^2} = \frac{1}{2} \left(\frac{w_{cf}}{w_{ga}} \right)^2 = \frac{1}{2\pi N_{ga}} \quad (3)$$

This expression also defines a Fresnel number for the gaussian aperture as $N_{ga} \equiv w_{ga}^2/2ab\lambda = \pi^{-1}(w_{ga}/w_{cf})^2$. In physical terms the quantity πN_{ga} is thus the ratio of the gaussian aperture area πw_{ga}^2 to the minimum possible stable mode area πw_{cf}^2 occurring at the same aperture plane in the limiting case of a confocal resonator, which has $m = 0$. In practical situations δ is usually small and N_{ga} large compared to unity.

The complex gaussian \tilde{q} parameter for the self-consistent Hermite-gaussian eigenmodes on the left-hand end mirror surface will then be given in general by [6]

$$\frac{1}{\tilde{q}} = \frac{1}{\tilde{B}} \left[- \left(\frac{\tilde{A} - \tilde{D}}{2} \right) \mp \sqrt{\left(\frac{\tilde{A} + \tilde{D}}{2} \right)^2 - 1} \right] = \mp \frac{m+1}{B} \sqrt{\frac{\tilde{m}-1}{\tilde{m}+1}} \quad (4)$$

The corresponding eigenvalue for the one-dimensional Hermite-gaussian mode $\tilde{u}_n(x)$

of order n will be

$$\tilde{\gamma}_n = \left(\frac{1}{\tilde{A} + \tilde{B}/\tilde{q}} \right)^{n+1/2} = \left(\frac{1}{\tilde{m} \mp \sqrt{\tilde{m}^2 - 1}} \right)^{n+1/2} \quad (5)$$

or, in two transverse dimensions, where the mode is $\tilde{u}_{np}(x, y)$, it will be

$$\tilde{\gamma}_{np} = \left(\frac{1}{\tilde{m} \mp \sqrt{\tilde{m}^2 - 1}} \right)^{n+p+1} \quad (6)$$

for the Hermite-Gaussian mode of indices (n, p) . One must select matching upper or lower signs in both Equations (4) and (6) so as to make $\text{Im}[1/\tilde{q}] < 0$, corresponding to a confined gaussian beam with $1/\tilde{q} \equiv 1/R - j\lambda/\pi w^2$, and also to make $|\tilde{\gamma}_{np}| < 1$.

B. Excess Noise Factor

For the lowest-order gaussian mode with gaussian parameter $1/\tilde{q} \equiv 1/R - j\lambda/\pi w^2$, the normalized transverse eigenfunction \tilde{u}_{00} on the left-hand mirror is (in two transverse dimensions, with $r^2 \equiv x^2 + y^2$)

$$\tilde{u}_{00}(x, y) = \left(\frac{2}{\pi w^2} \right)^{1/2} \exp \left(-j \frac{\pi r^2}{\tilde{q} \lambda} \right) \quad (7)$$

where w is determined from the imaginary part of $1/\tilde{q}$, and where the eigenfunction is power-normalized so that $\int \tilde{u}_{00}^*(\mathbf{s}) \tilde{u}_{00}(\mathbf{s}) d\mathbf{s} = 1$, with $\mathbf{s} \equiv (x, y)$, $d\mathbf{s} \equiv dx dy$, and the integration taken over the whole transverse plane. Since this function is determined at a symmetry plane where $\tilde{u}(\mathbf{s})$ and $\tilde{\phi}(\mathbf{s})$ have the same form (using the notation from References [3] and [4]), the biorthogonality condition $\int \tilde{u}_{00}(\mathbf{s}) \tilde{\phi}_{00}(\mathbf{s}) d\mathbf{s} = 1$ requires the adjoint function $\tilde{\phi}_{00}$ to have the normalized form

$$\tilde{\phi}_{00}(x, y) = j \frac{(2\pi)^{1/2} w}{\tilde{q} \lambda} \exp \left(-j \frac{\pi r^2}{\tilde{q} \lambda} \right) \quad (8)$$

The eigenfunction \tilde{u}_{00} and the adjoint function $\tilde{\phi}_{00}$ have the same form at this plane, but must have different normalizations. The excess noise factor for this mode is then given in two transverse dimensions, using the results of [3] and [4], by

$$\begin{aligned} K_{00} &\equiv \int_0^\infty 2\pi r \tilde{\phi}_{00}(r) \tilde{\phi}_{00}^*(r) dr \\ &= \frac{4\tilde{q}\tilde{q}^*}{[\tilde{q}^* - \tilde{q}]^2} = 1 + (\pi w^2/R\lambda)^2 \end{aligned} \quad (9)$$

Finite wavefront curvature R at the symmetry plane is obviously a measure both of the degree of instability of the resonator mode and the degree to which the excess noise factor exceeds unity. Since we have \tilde{q} and \tilde{q}^* in terms of m and N_{ga} , the excess noise factor K_{00} is determined as a function of m and N_{ga} .

Values of excess noise factors and of the so-called “mode cross-powers” for the higher-order Hermite-Gaussian transverse modes can also be determined. The analytical expressions for these higher-order modes become quite complicated, and are briefly summarized in the Appendix.

III. RESULTS AND DISCUSSION

The basic physical significance of the Petermann noise factor K_{np} , interpreted as an excess noise factor for the np -th transverse mode of a laser cavity, is extensively discussed in references [3] and [4]. In brief, the spontaneous atomic emission rate into such a cavity mode appears to be enhanced by a factor of K_{np} , as if there were K_{np} excess noise photons per mode instead of the usual 1 noise photon per mode. The quantum noise fluctuations, or Schawlow-Townes fluctuations, in the corresponding laser oscillator will be enhanced over usual values by the same factor. Our purpose here is to show in particular how large this noise enhancement can become for geometrically unstable VRM laser cavities.

Figures 2 and 3 show representative values of both the power loss per round trip $1-|\tilde{\gamma}_{00}|^2$ calculated from Equation (6) and also the excess noise factor K_{00} calculated from Equation (9) plotted against geometrical magnification M and aperture Fresnel number N_{ga} in typical VRM cavities. Note that for $m > 1$, $M \equiv m + \sqrt{m^2 - 1}$ is the geometrical magnification of the unstable cavity. The horizontal scale variable in fact is the half-trace parameter m itself for $0 < m < 1$, shifting to the geometrical magnification M for $m > 1$. The rapid increase of the excess noise factor with both of these parameters for $m \geq 1$ is evident. Figures 4 to 8 present similar results obtained for higher-order Hermite-gaussian modes. Both the excess noise factor and the power loss per round trip obviously increase rapidly with the order of the mode.

Useful analytical approximations making use of the generally small value of $\delta \equiv 1/2\pi N_{ga}$ can also be developed for various special cases of interest, as follows:

1) *Strongly unstable positive-branch VRM resonators* (m and $M > 1$): The resonator eigenvalue in this case is to a good approximation given simply by $\tilde{\gamma}_{np} = (1/M)^{n+p+1}$, independent of the gaussian aperture parameter δ . In essence, as one changes the aperture size in the cavity at fixed magnification, the mode size in the cavity changes in direct proportion to w_{ga} . The excess noise factor for the gaussian lowest-order mode is given in this case to a good approximation by

$$K_{00} \approx 1 + \left(\frac{m^2 - 1}{\delta} \right)^2 \approx \left[\frac{\pi}{2} \left(\frac{M^2 - 1}{M} \right)^2 N_{ga} \right]^2 \quad (10)$$

so that K_{00} rises proportional to both M^4 and N_{ga}^2 at large enough values. For

higher-order modes, this result generalizes to

$$K_{np} \approx C_{np} \left(\frac{m^2 - 1}{\delta} \right)^{2(n+p+1)} \approx C_{np} \left[\frac{\pi}{2} \left(\frac{M^2 - 1}{M} \right)^2 N_{ga} \right]^{2(n+p+1)} \quad (11)$$

with a proportionality constant C_{np} given by

$$C_{np} = \left(\frac{(2n)! (2p)!}{2^{n+p} (n!)^2 (p!)^2} \right)^2 \quad (12)$$

2) *Planar VRM resonators* (m and $M = 1$): A planar resonator, with $M = m = 1$, marks the boundary between stable and unstable regions. For $m = 1$ and $\delta \ll 1$ the resonator eigenvalue becomes

$$\tilde{\gamma}_{np} \approx \left(1 - (1 - j)\sqrt{\delta} \right)^{n+p+1} \quad (13)$$

and hence the power loss per round trip is

$$1 - |\tilde{\gamma}_{np}|^2 \approx 2(n + p + 1) \sqrt{\delta} \quad (14)$$

The excess noise factor for the gaussian lowest-order mode correspondingly becomes

$$K_{00} \approx 2/[1 - \delta/2] \approx 2 \quad (15)$$

for $\delta \ll 1$. The excess noise is to a good approximation independent of δ for higher-order modes.

3) *Geometrically stable VRM resonators* ($|m| < 1$): So long as the gaussian aperture is small enough so that δ is small compared to $(1 - m^2)$, the eigenvalue for a geometrically stable VRM resonator can be written as

$$\tilde{\gamma}_{np} \approx [e^{j\theta}(1 - \delta/\sin \theta)]^{n+p+1} \quad (16)$$

where $\cos \theta \equiv m$. The power loss per round trip in this case, namely

$$1 - |\tilde{\gamma}_{np}|^2 \approx 2(n + p + 1) \delta / \sqrt{1 - m^2}, \quad (17)$$

is then small and the excess noise factor is given by

$$K_{np} \approx 1 + (n + p + 1) [\delta / (1 - m^2)]^2 \quad (18)$$

Hence, K_{np} is very close to unity over nearly all the stable range until $|m| \rightarrow 1$.

4) *Small aperture limit*, $w_{ga}^2 \rightarrow 0$: Finally, if the aperture becomes very small, so that $N_{ga} \rightarrow 0$ or $\delta \rightarrow \infty$, independent of the half-trace parameter m , the excess noise factor blows up as

$$K_{np} \approx C_{np} \delta^{2(n+p+1)} \approx C_{np} \left(\frac{1}{2\pi N_{ga}} \right)^{2(n+p+1)} \quad (19)$$

independent of m . The gaussian mode in this case spreads very rapidly, going out from the very small gaussian aperture so that the value of m , or the size of the mode striking the aperture, does not matter in determining K_{np} . The eigenvalue is given by

$$\tilde{\gamma}_{np} \approx (j/2\delta)^{n+p+1} \quad (20)$$

i.e., the cavity loss per bounce becomes very large.

For Laguerre-gaussian modes of indices (n, ℓ) , similar expressions can be obtained by simply replacing the quantity $n + p + 1$ by $2n + |\ell| + 1$, and the proportionality constant C_{np} by

$$C_{n\ell} = \left[\binom{2n + |\ell|}{n} \right]^2 = \left[\frac{(2n + |\ell|)!}{n! (n + |\ell|)!} \right]^2 \quad (21)$$

where n is the usual radial index and ℓ the azimuthal index for the eigenmode functions $\tilde{u}_{n\ell}(r, \theta)$.

Perhaps the most surprising aspect of all these results is how large the excess emission factor K_{np} can become for quite reasonable values of M and N_{ga} . Experiments to test these calculations in real lasers would be very desirable and are now being initiated.

APPENDIX

The analytical overlap integrals used in the preceding analysis can become fairly complex for higher-order gaussian modes. Therefore, in this appendix, we summarize for the potential use of other workers the general analytical results obtained for the excess noise factors and cross-power coefficients of higher-order modes, both in cartesian (Hermite-gaussian modes) and cylindrical (Laguerre-gaussian modes) coordinate systems.

A. Cartesian coordinates: Hermite-gaussian modes

1) *Transverse eigenmodes (Forward-propagating modes)*: The eigenmodes $\tilde{u}_{np}(x, y)$ of the complex Huygens' integral equation in cartesian coordinates are the complex Hermite-gaussian modes defined as

$$\tilde{u}_{np}(x, y) = \tilde{\alpha}_{np} \tilde{v}^{n+p} H_n\left(\frac{\sqrt{2}}{\tilde{v}}x\right) H_p\left(\frac{\sqrt{2}}{\tilde{v}}y\right) \exp\left\{-\frac{j\pi}{\lambda\tilde{q}}(x^2 + y^2)\right\} \quad (22)$$

with $H_n(z)$ the complex n^{th} -order Hermite polynomial, $1/\tilde{q} \equiv 1/R - j\lambda/\pi w^2$ the complex gaussian parameter and \tilde{v} a complex scaling parameter. These self-reproducing parameters are given by Equation (4) and by $1/\tilde{v}^2 = j\pi/\lambda\tilde{q}$. The eigenfunctions are power-normalized by the condition $\int \tilde{u}_{np}^*(\mathbf{s}) \tilde{u}_{np}(\mathbf{s}) d\mathbf{s} = 1$ which gives as normalization coefficient

$$|\tilde{\alpha}_{np}|^2 = |\tilde{v}|^{-2(n+p)} I_n^{-1} I_p^{-1} \quad (23)$$

In this expression, we have introduced $I_n = \tilde{I}_{nn}$, where we define the function $\tilde{I}_{n_1 n_2}$ as

$$\tilde{I}_{n_1 n_2} = w \sqrt{\frac{\pi}{2}} \sum_{i=0}^{i_m} n_1! n_2! \frac{(2|\tilde{\eta}|^2)^i}{i!} \frac{(\tilde{\eta}^2 - 1)^{(n_1-i)/2}}{\left(\frac{n_1-i}{2}\right)!} \frac{(\tilde{\eta}^{*2} - 1)^{(n_2-i)/2}}{\left(\frac{n_2-i}{2}\right)!} \quad (24)$$

with $\tilde{\eta} = w/\tilde{v}$, $i_m = \min(n_1, n_2)$, and the summation being limited to values of i such that $(n_1 - i)/2$ and $(n_2 - i)/2$ be integer. This expression can be reduced to

$$\begin{aligned} \tilde{I}_{n_1 n_2} &= w \sqrt{\frac{\pi}{2}} \frac{n_1!}{(n_1/2)!} \frac{n_2!}{(n_2/2)!} (\tilde{\eta}^2 - 1)^{n_1/2} (\tilde{\eta}^{*2} - 1)^{n_2/2} \\ &\quad \times {}_2F_1 \left(-\frac{n_1}{2}, -\frac{n_2}{2}; \frac{1}{2}; \left[\frac{|\tilde{\eta}|^2}{|\tilde{\eta}^2 - 1|} \right]^2 \right) \quad \text{for } n_1, n_2 \text{ both even} \end{aligned} \quad (25)$$

$$\begin{aligned} &= w \sqrt{\frac{\pi}{2}} \frac{n_1!}{((n_1 - 1)/2)!} \frac{n_2!}{((n_2 - 1)/2)!} (\tilde{\eta}^2 - 1)^{(n_1 - 1)/2} (\tilde{\eta}^{*2} - 1)^{(n_2 - 1)/2} 2|\tilde{\eta}|^2 \\ &\quad \times {}_2F_1 \left(-\frac{n_1 - 1}{2}, -\frac{n_2 - 1}{2}; \frac{3}{2}; \left[\frac{|\tilde{\eta}|^2}{|\tilde{\eta}^2 - 1|} \right]^2 \right) \quad \text{for } n_1, n_2 \text{ both odd} \quad (26) \\ &= 0 \quad \text{for } n_1, n_2 \text{ having different parities} \end{aligned}$$

where ${}_2F_1(a, b; c; z)$ is a hypergeometric function. For $|\tilde{\eta}| \gg 1$, we have $[|\tilde{\eta}|^2/|\tilde{\eta}^2 - 1|]^2 \rightarrow 1$, leading to the asymptotic formulas of Equations (11) & (19), where we used the property ${}_2F_1(a, b; c; 1) = \Gamma(c)\Gamma(c - a - b)/\Gamma(c - a)\Gamma(c - b)$.

2) *Adjoint transverse eigenmodes (Backward-propagating modes)*: The adjoint eigenmodes $\tilde{\phi}_{np}$ are given by

$$\tilde{\phi}_{np}(x, y) = \tilde{\beta}_{np} \tilde{v}^{n+p} H_n \left(\frac{\sqrt{2}}{\tilde{v}} x \right) H_p \left(\frac{\sqrt{2}}{\tilde{v}} y \right) \exp \left\{ \left[\frac{j\pi}{\lambda \tilde{q}} - \frac{2}{\tilde{v}^2} \right] (x^2 + y^2) \right\} \quad (27)$$

These adjoint functions are normalized by the biorthogonality condition $\int \tilde{u}_{np}(\mathbf{s}) \tilde{\phi}_{np}(\mathbf{s}) d\mathbf{s} = 1$ which gives as normalization coefficient

$$\tilde{\beta}_{np} = \frac{2}{\pi} \left(2^{n+p} n! p! \tilde{\alpha}_{np} \tilde{v}^{2(n+p+1)} \right)^{-1} \quad (28)$$

3) *Excess noise factor*: The excess noise factor K_{np} for the Hermite-gaussian

mode of indices (n, p) can be expressed as

$$K_{np} = \int \tilde{\phi}_{np}^*(\mathbf{s}) \tilde{\phi}_{np}(\mathbf{s}) d\mathbf{s} \quad (29)$$

$$= |\tilde{\beta}_{np}|^2 |\tilde{v}|^{2(n+p)} I_n I_p \quad (30)$$

$$= \left(\frac{2}{\pi}\right)^2 \left(2^{2(n+p)} (n!)^2 (p!)^2 |\tilde{v}|^4\right)^{-1} I_n^2 I_p^2 \quad (31)$$

4) *Cross-power*: The cross-power coefficient $\tilde{P}_{n_1 n_2 p_1 p_2}$ between the Hermite-gaussian modes of indices (n_1, p_1) and (n_2, p_2) can be expressed as

$$\tilde{P}_{n_1 n_2 p_1 p_2} = \int \tilde{u}_{n_1 p_1}^*(\mathbf{s}) \tilde{u}_{n_2 p_2}(\mathbf{s}) d\mathbf{s} \quad (32)$$

$$= \tilde{\alpha}_{n_1 p_1} \tilde{\alpha}_{n_2 p_2}^* \tilde{v}^{n_1+p_1} \tilde{v}^{* n_2+p_2} \tilde{I}_{n_1 n_2} \tilde{I}_{p_1 p_2} \quad (33)$$

Note that the cross-power coefficient is zero if n_1 and n_2 , or p_1 and p_2 , have different parities.

B. Cylindrical coordinates: Laguerre-gaussian modes

1) *Transverse eigenmodes (Forward-propagating modes)*: The eigenmodes $\tilde{u}_{n\ell}(r, \theta)$ of the complex Huygens' integral equation in cylindrical coordinates are the complex Laguerre-gaussian modes defined as

$$\tilde{u}_{n\ell}(r, \theta) = \tilde{a}_{n\ell} \left(\frac{\sqrt{2}}{\tilde{v}} r\right)^{|\ell|} L_n^\ell \left(\frac{2}{\tilde{v}^2} r^2\right) \exp \left\{ -\frac{j\pi}{\lambda \tilde{q}} r^2 + j\ell\theta \right\} \quad (34)$$

with $L_n^\ell(z)$ the complex generalized Laguerre polynomial, and the parameters \tilde{v} and \tilde{q} defined as in cartesian coordinates. The power-normalization gives here as normalization coefficient

$$|\tilde{a}_{n\ell}|^2 = \frac{1}{2\pi} \left(\frac{\sqrt{2}}{|\tilde{v}|}\right)^{-2|\ell|} M_{n\ell}^{-1} \quad (35)$$

In this expression, we have introduced $M_{n\ell} = \tilde{M}_{nn\ell\ell}$, where we define the function $\tilde{M}_{n_1 n_2 \ell_1 \ell_2}$ as

$$\begin{aligned} \tilde{M}_{n_1 n_2 \ell_1 \ell_2} = \delta_{\ell_1 \ell_2} \frac{1}{2} \left(\frac{w^2}{2} \right)^{|\ell_1|+1} \sum_{i=0}^{n_1} \sum_{k=0}^{n_2} \left[\binom{n_1 + |\ell_1|}{n_1 - i} \binom{n_2 + |\ell_1|}{n_2 - k} \right. \\ \left. \times \frac{1}{i!k!} \left(\frac{-w^2}{\tilde{v}^2} \right)^i \left(\frac{-w^2}{\tilde{v}^{*2}} \right)^k (i+k+|\ell_1|)! \right] \end{aligned} \quad (36)$$

2) *Adjoint transverse eigenmodes (Backward-propagating modes)*: The adjoint eigenmodes $\tilde{\phi}_{n\ell}$ in cylindrical coordinates are given by

$$\tilde{\phi}_{n\ell}(r, \theta) = \tilde{b}_{n\ell} \left(\frac{\sqrt{2}}{\tilde{v}} r \right)^{|\ell|} L_n^\ell \left(\frac{2}{\tilde{v}^2} r^2 \right) \exp \left\{ \left[\frac{j\pi}{\lambda \tilde{q}} - \frac{2}{\tilde{v}^2} \right] r^2 - i\ell\theta \right\} \quad (37)$$

The biorthogonality condition gives as normalization coefficient

$$\tilde{b}_{n\ell} = \frac{2}{\pi} \frac{n!}{(n+|\ell|)!} \frac{1}{\tilde{a}_{n\ell} \tilde{v}^2} \quad (38)$$

3) *Excess noise factor*: The excess noise factor $K_{n\ell}$ for the Laguerre-gaussian mode of indices (n, ℓ) can be expressed as

$$K_{n\ell} = |\tilde{b}_{n\ell}|^2 2\pi \left(\frac{\sqrt{2}}{|\tilde{v}|} \right)^{2|\ell|} M_{n\ell} \quad (39)$$

$$= 4 \left(\frac{2}{\pi} \frac{n!}{(n+|\ell|)!} \right)^2 \left(\frac{2}{|\tilde{v}|^2} \right)^{2(|\ell|+1)} M_{n\ell}^2 \quad (40)$$

4) *Cross-power*: The cross-power coefficient $\tilde{P}_{n_1 n_2 \ell_1 \ell_2}$ between the Laguerre-gaussian modes of indices (n_1, ℓ_1) and (n_2, ℓ_2) can be expressed as

$$\tilde{P}_{n_1 n_2 \ell_1 \ell_2} = \delta_{\ell_1 \ell_2} 2\pi \tilde{a}_{n_1 \ell_1} \tilde{a}_{n_2 \ell_2}^* \left(\frac{\sqrt{2}}{|\tilde{v}|} \right)^{2|\ell_1|} \tilde{M}_{n_1 n_2 \ell_1 \ell_2} \quad (41)$$

ACKNOWLEDGEMENTS

This work was carried out with support from the U.S. Air Force Office of Scientific Research. One of us (AES) is also very grateful for support from the Alexander von Humbolt Foundation and the hospitality of the Max-Planck Institut für Quantenoptik in Garching, West Germany. Two of us (JLD & PLM) are grateful for support from the Belgian American Educational Foundation.

REFERENCES

- [1] K. Petermann. "Calculated spontaneous emission factor for double-heterostructure injection lasers with gain-induced waveguiding," *IEEE J. Quantum Electron.*, vol. QE-15, pp. 566-570, July 1979.
- [2] H. A. Haus and S. Kawakami, "On the "excess spontaneous emission factor" in gain-guided laser amplifiers," *IEEE J. Quantum Electron.*, vol. QE-21, pp. 63-69, January 1985.
- [3] A. E. Siegman. "Excess spontaneous emission in nonhermitian optical systems. I. Laser amplifiers," *Phys. Rev. A*, vol. 39, pp. 1253-1263, February 1989.
- [4] A. E. Siegman. "Excess spontaneous emission in nonhermitian optical systems. II. Laser oscillators," *Phys. Rev. A*, vol. 39, pp. 1264-1268, February 1989.
- [5] M. A. Lauder and G. H. New, "Biorthogonality properties and excess noise factors of unstable optical resonators," *Optics Communications*, vol. 67, pp. 343-348, August 1988.
- [6] A. E. Siegman. *Lasers*, University Science Books, Mill Valley, CA, 1986.

Jean-luc Doumont was born in Brussels, Belgium, on April 20, 1964. He received the degree of Engineer in Applied Physics from the Université Catholique de Louvain (Catholic University of Louvain), Louvain-la-Neuve, Belgium in 1987. He has been a Technical Student at CERN (European Organization for Nuclear Research) in 1987, where he worked on laser calibration of drift chambers.

He is currently working towards the Ph.D. degree in Applied Physics at Stanford University, Stanford, California. His research interests include laser physics and the study of laser resonators.

Jean-luc Doumont is a fellow from the Belgian American Educational Foundation.

Paul L. Mussche was born in Nottingham, Great Britain, on October 20, 1962. He received a degree in Electrical Engineering from the Université Libre de Bruxelles (Free University of Brussels), Brussels, Belgium in 1985. His engineering thesis was concerned with optical computing. He then came to Stanford University, Stanford, California and earned a Master of Science in Electrical Engineering in 1986.

He is currently working towards his Ph.D. in Electrical Engineering at Stanford. His research interests include laser physics and in particular quantum noise in lasers and optical resonator mode calculations.

Paul L. Mussche is a member of the Optical Society of America and a fellow of the Belgian American Educational Foundation.

Anthony E. Siegman holds the Burton J. and Ann M. McMurtry Professorship of Engineering in the Departments of Electrical Engineering and (by courtesy) of Applied Physics and is currently (1988-89) also Director of the Edward L. Ginzton Laboratory at Stanford University.

Professor Siegman was born in Detroit on November 23, 1931. He received the A.B. degree from Harvard in 1952; the M.S. in Applied Physics from UCLA in 1954 under the Hughes Aircraft Cooperative Plan; and the Ph.D. in Electrical Engineering from Stanford in 1957. Since then he has been on the faculty at Stanford, where he directs an active research program in lasers and their applications.

Professor Siegman has made many contributions to the fields of microwave electronics, traveling wave tubes, laser devices and lasers applications, and has written a highly regarded textbook on Lasers (University Science Books, 1986). He has been a consultant to numerous industrial and government laboratories, and served for six years on the Air Force Scientific Advisory Board.

Professor Siegman is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Optical Society of America, and the American Physical Society, and a member of the National Academies of Engineering (1973) and of Science (1988). In 1980 he received the R.W. Wood Prize of the Optical Society of America for the invention of the unstable optical resonator, and in 1987 was awarded the Frederic Ives Medal of the Optical Society of America for overall distinction in optics.

CAPTIONS FOR FIGURES

Figure 1. Analytical model for a gaussian VRM laser cavity containing arbitrary real optical elements with an overall ray matrix $[abcd]$ going to the right from a 100% reflecting mirror surface on the left-hand end to a gaussian variable-reflectivity mirror on the right end: (1a) standing wave cavity ; (1b) lensguide equivalent.

Figure 2. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{00}|^2$ in percent per round trip and excess noise factor K_{00} plotted versus geometrical magnification for the lowest-order Hermite-gaussian mode in a VRM cavity having a ratio of gaussian aperture to confocal spot size of $w_{ga}/w_{cf} = 5$ or gaussian Fresnel number $N_{ga} \approx 8$. The horizontal scale variable is the half-trace parameter m for $0 < m < 1$, and the geometrical magnification $M \equiv m + \sqrt{m^2 - 1}$ for $m > 1$.

Figure 3. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{00}|^2$ per round trip (lower plot) and excess noise factor K_{00} (upper plot) plotted versus gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ for different values of the geometrical magnification M in a gaussian VRM cavity operating in the lowest-order Hermite-gaussian mode.

Figure 4. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{np}|^2$ per round trip (lower plot) and excess noise factor K_{np} (upper plot) plotted versus geometrical magnification for a gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ of 1 and different values of the Hermite-gaussian mode indices (n, p) in a gaussian VRM cavity. The horizontal scale variable is the half-trace parameter m for $0 < m < 1$, and the geometrical magnification M for $m > 1$.

Figure 5. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{np}|^2$ per round trip (lower

plot) and excess noise factor K_{np} (upper plot) plotted versus gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ for a half-trace parameter m of 0.1 (stable gaussian VRM cavity) and different values of the Hermite-gaussian mode indices (n, p) .

Figure 6. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{np}|^2$ per round trip (lower plot) and excess noise factor K_{np} (upper plot) plotted versus gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ for a half-trace parameter $m = M = 1$ (planar gaussian VRM cavity) and different values of the Hermite-gaussian mode indices (n, p) .

Figure 7. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{np}|^2$ per round trip (lower plot) and excess noise factor K_{np} (upper plot) plotted versus gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ for a half-trace parameter M of 1.5 (unstable gaussian VRM cavity) and different values of the Hermite-gaussian mode indices (n, p) .

Figure 8. Power loss (or power output coupling) $1 - |\tilde{\gamma}_{np}|^2$ per round trip (lower plot) and excess noise factor K_{np} (upper plot) plotted versus gaussian aperture Fresnel number $N_{ga} \equiv \pi^{-1}(w_{ga}/w_{cf})^2$ for a half-trace parameter M of 2 (unstable gaussian VRM cavity) and different values of the Hermite-gaussian mode indices (n, p) .

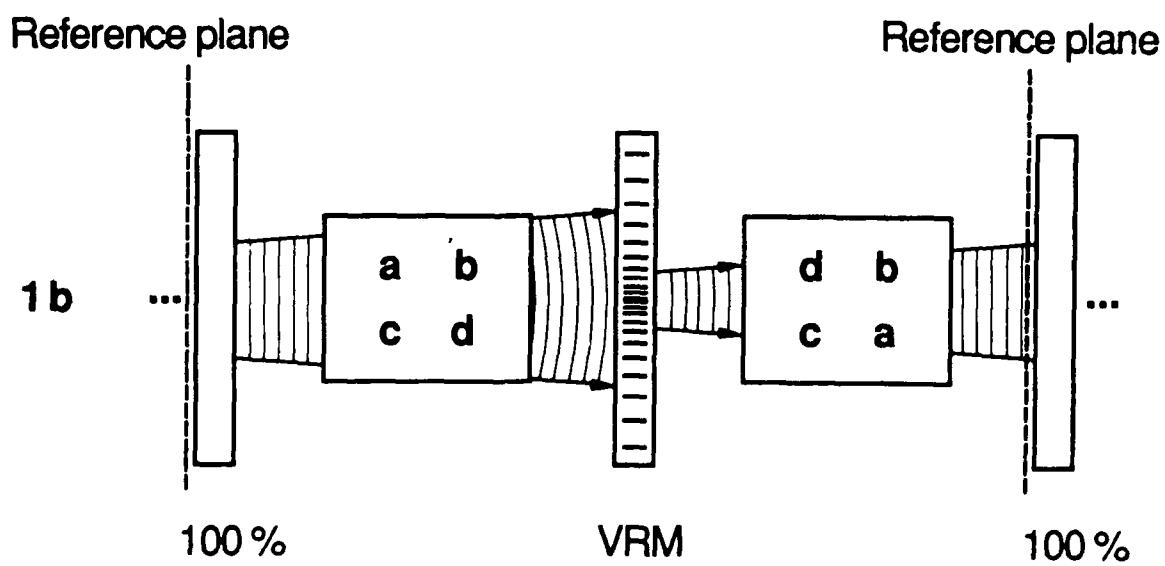
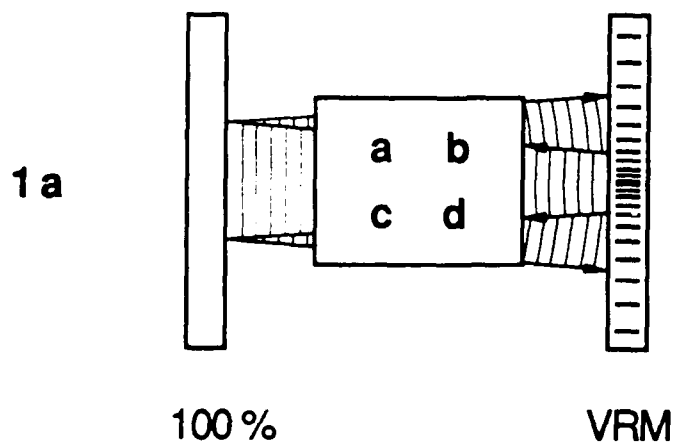


Fig. 1

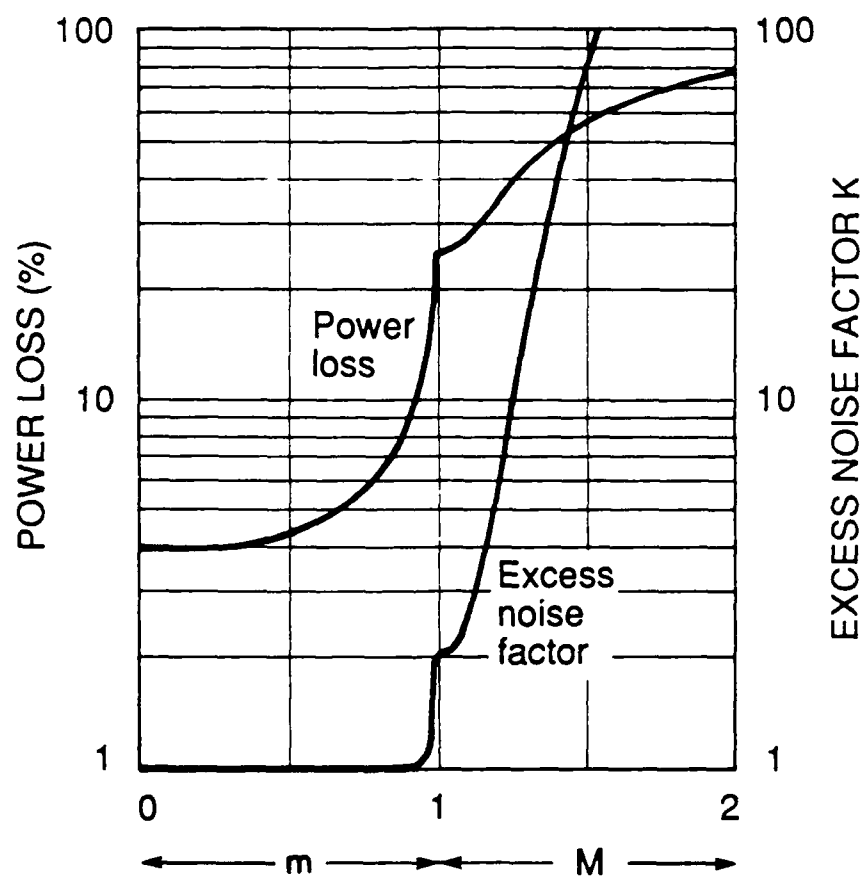


Fig. 2

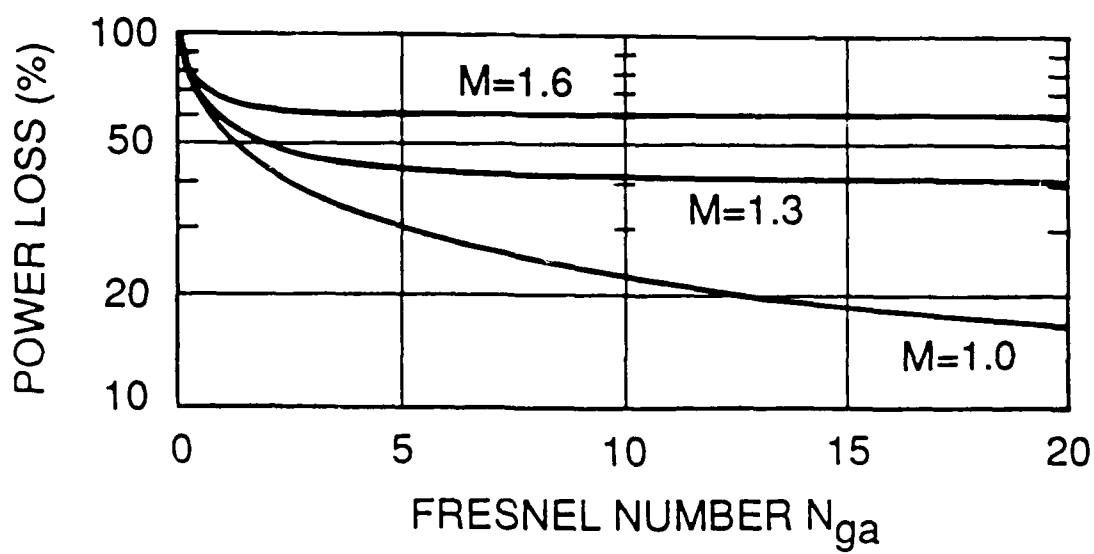
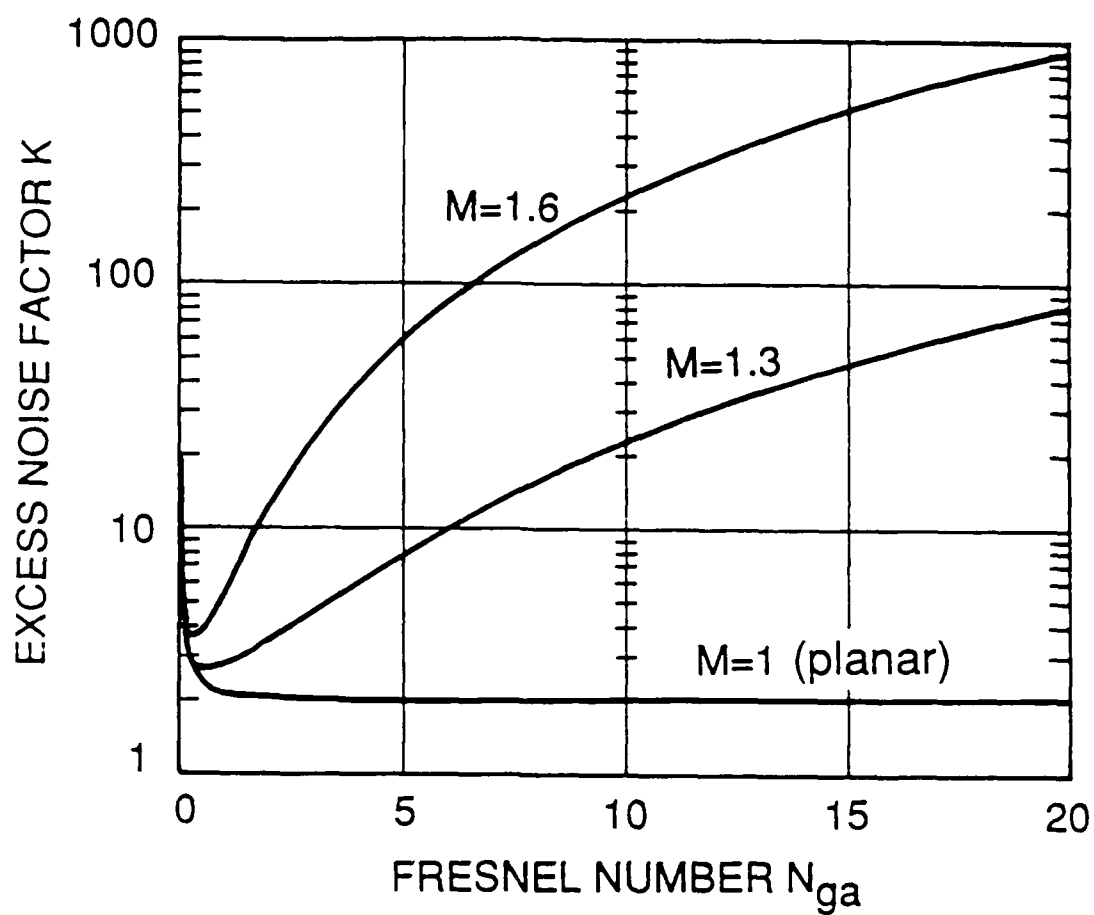


Fig. 3

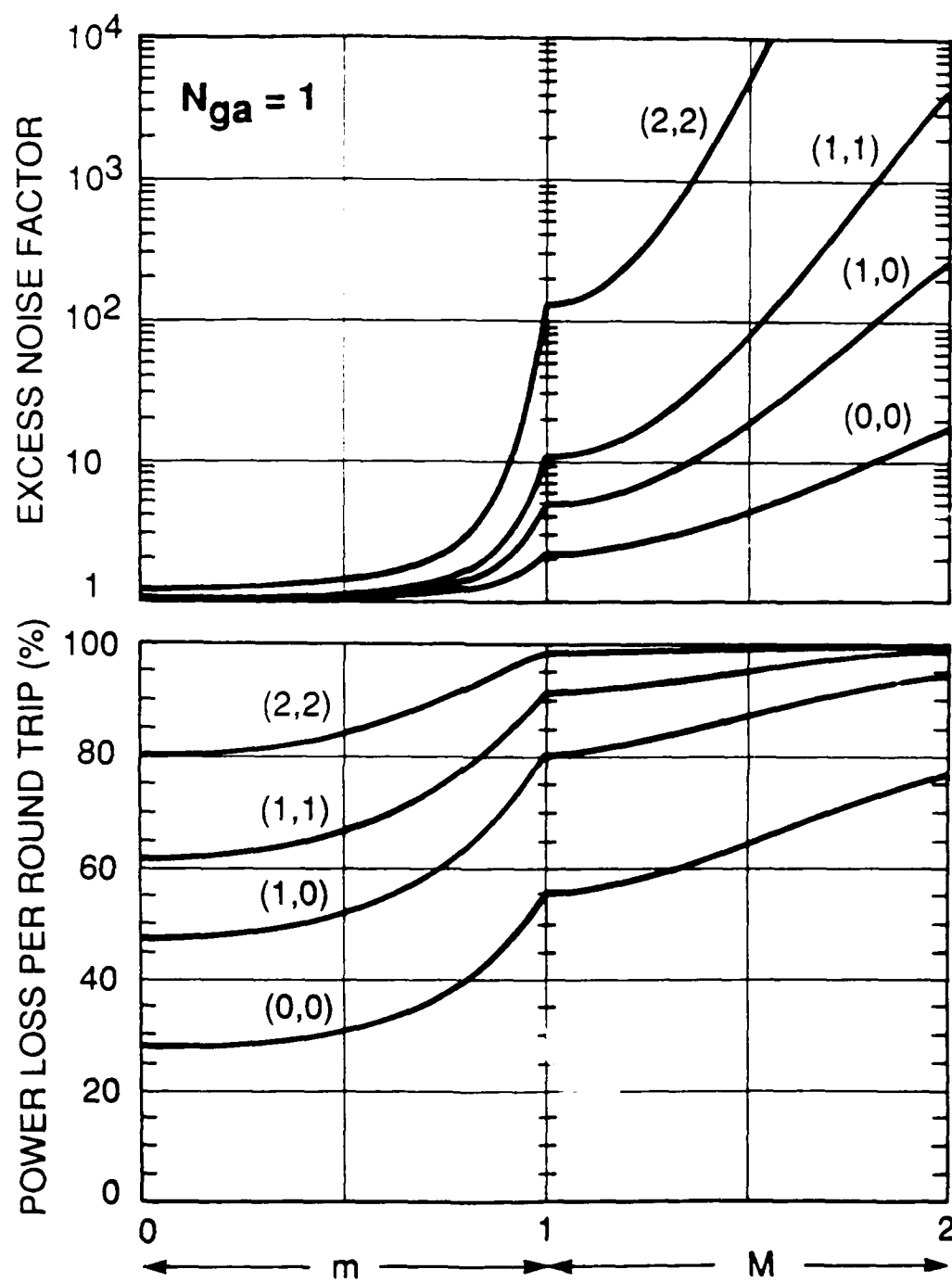


Fig. 4

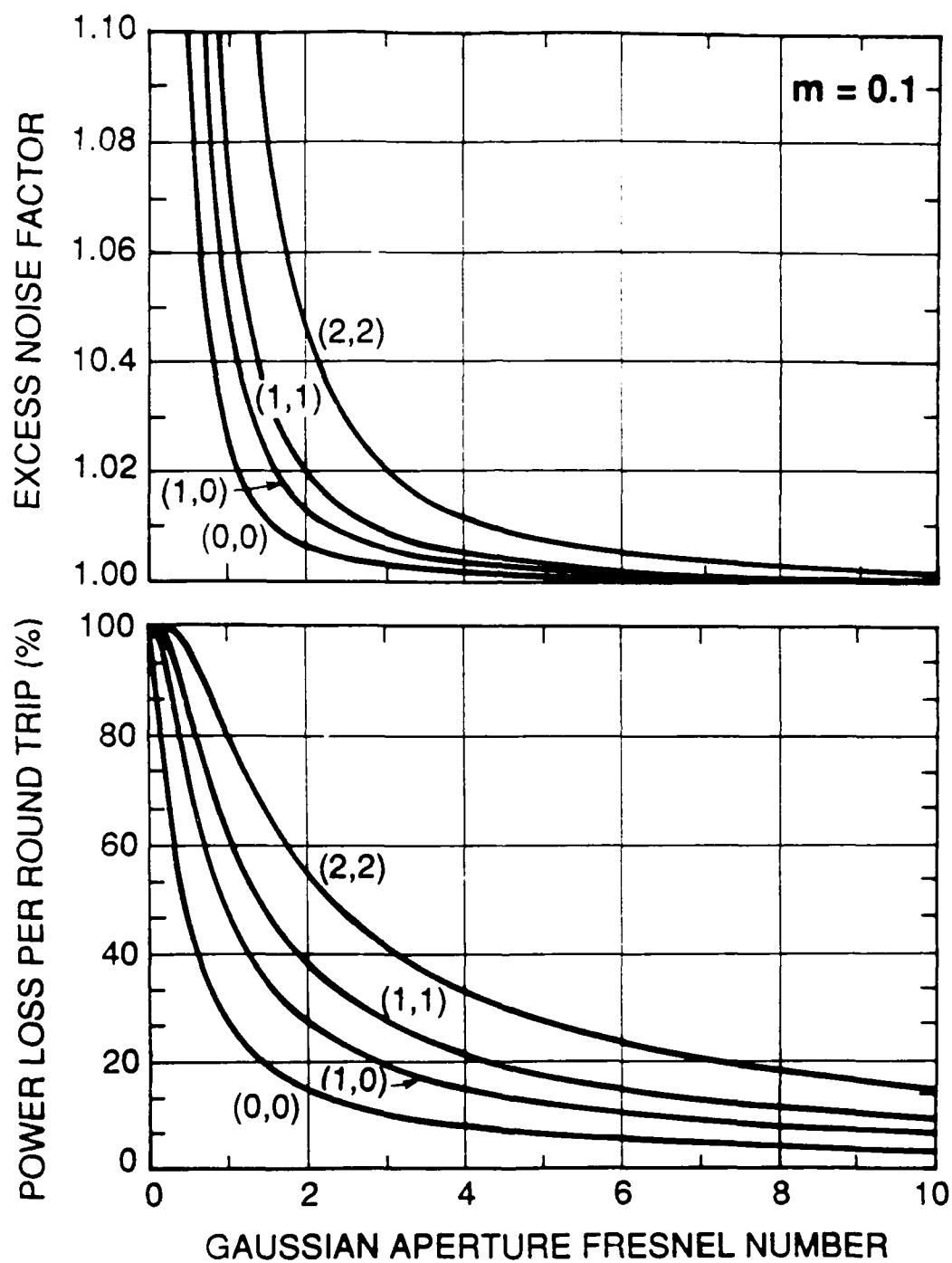


Fig. 5

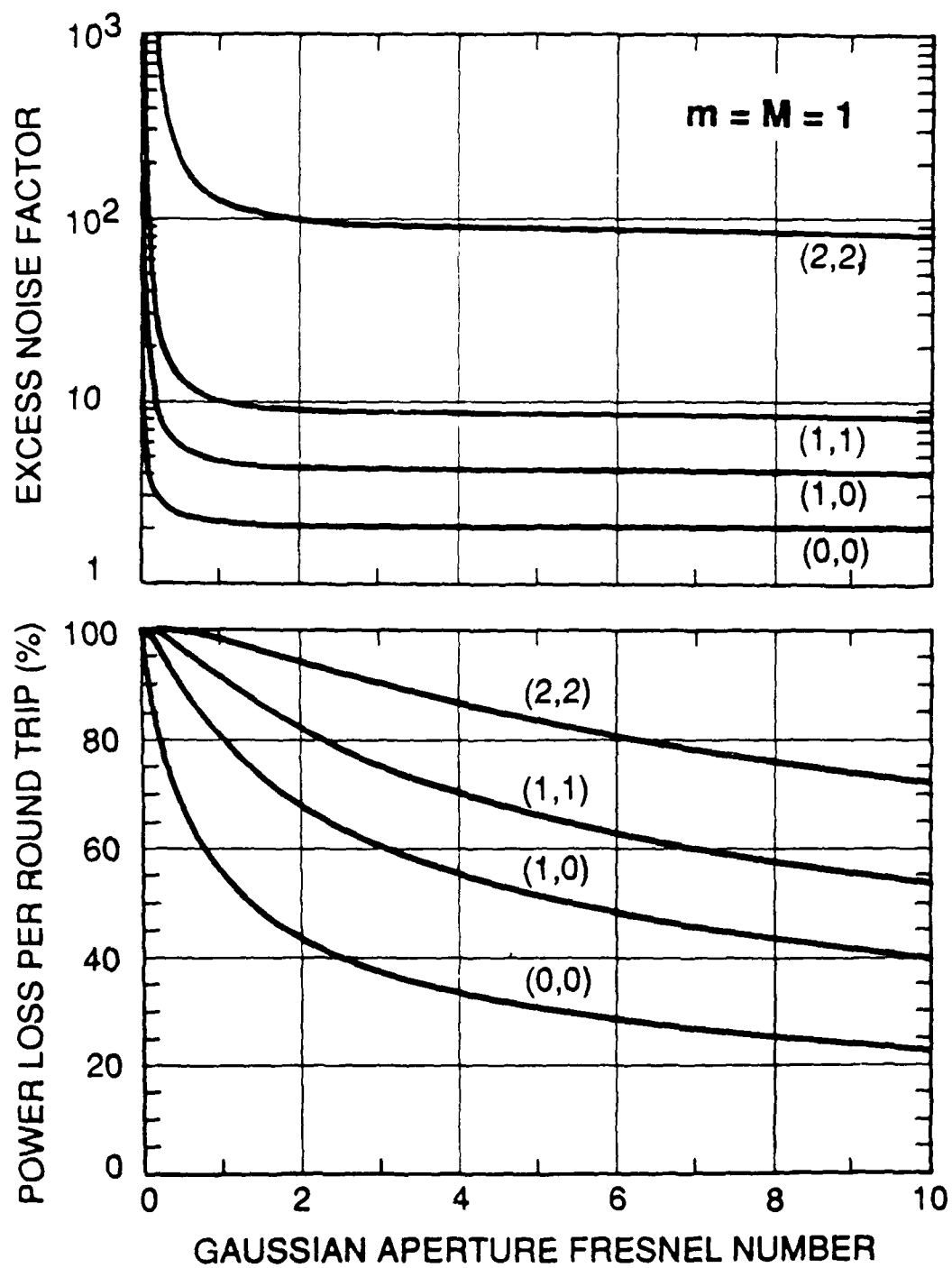


Fig. 6

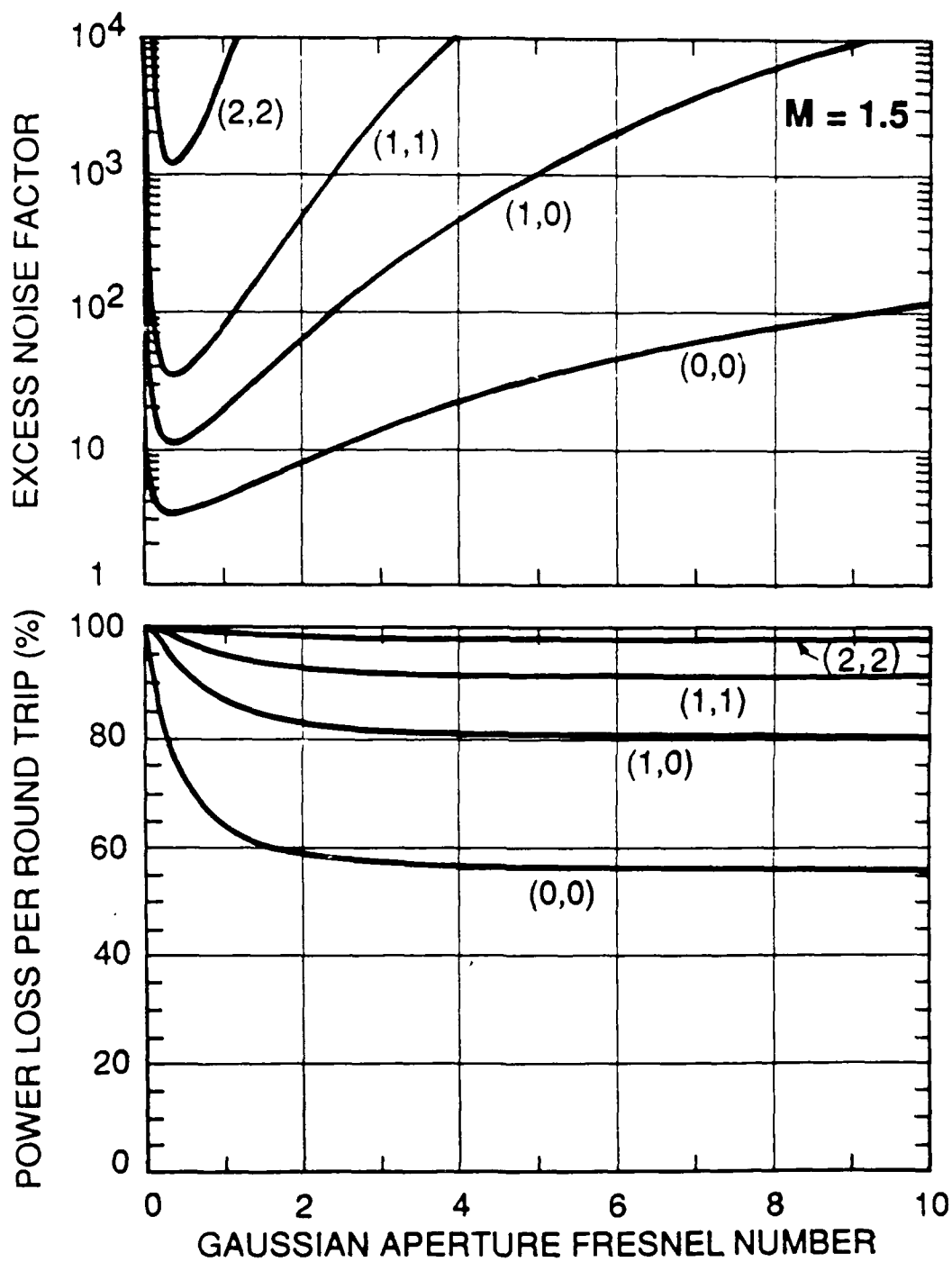


Fig. 7

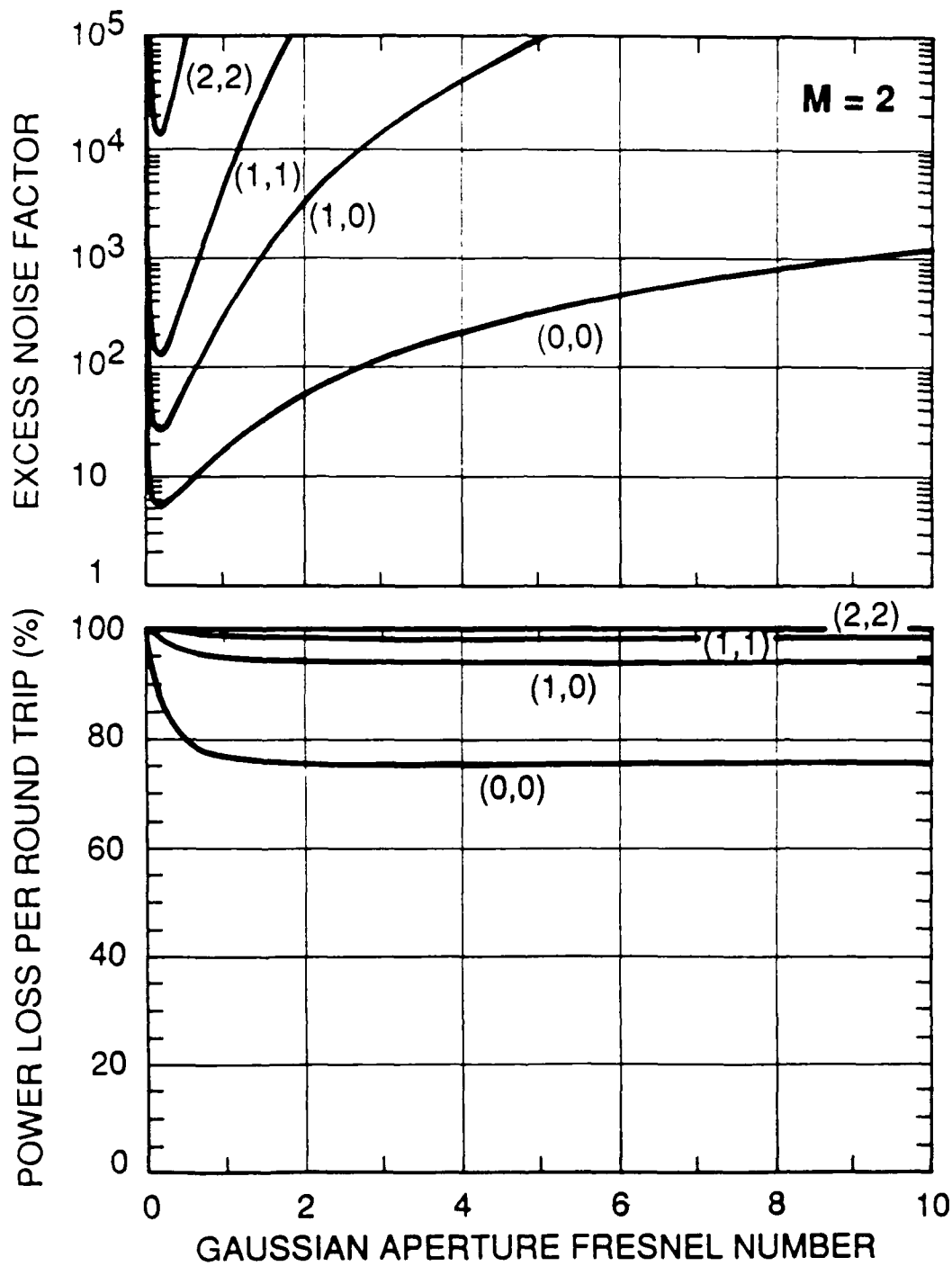


Fig. 8

Axial Modes in a Grating-Dispersed Laser Cavity

A. E. Siegman

Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA

Received 10 November 1986 Accepted 17 November 1986

Abstract. Two recent papers by Stankov and the author have predicted that unusual axial mode properties should occur in a laser cavity containing an intracavity dispersive grating pair. The analysis in these reports is unfortunately incorrect, and the corrected properties of the dispersive cavity are found to be less interesting than at first predicted.

PACS: 42.60

A recent publication in this journal by Stankov [1] and a postdeadline paper by Stankov and Siegman at the 1986 CLEO meeting [2] both described a scheme for greatly modifying the axial-mode properties of a laser cavity by including a dispersive grating pair within the cavity. Unfortunately, the analysis in these two reports was incorrect, and the resulting conclusions are erroneous. The purpose of this note is to acknowledge this error, and briefly summarize the correct (and unfortunately less interesting) behavior of the system.

To understand the error that was made in the analysis, it may be most effective to repeat it. The type of cavity under consideration can be modeled by the system shown in Fig. 1a. We think of a slender well-collimated laser beam bouncing back and forth between the end mirrors M_1 and M_2 along the path shown in this figure. In following this path the optical beam is diffracted twice by each of the two identical transmission gratings G_1 and G_2 in each round trip. The beam path is pinned in transverse position by an aperture at the G_1 end of the cavity, and intersects grating G_2 at a transverse position $x=a$ which depends on the diffraction angle θ and hence the optical frequency ω .

For the geometry shown in Fig. 1a, the wavelength-dependent path length $L(\omega)$ between the intersection points on the two gratings will be given by

$$L(\omega) = \frac{L_0}{\cos \theta} = \frac{L_0}{\sqrt{1 - (k_g/k)^2}}, \quad (1)$$

where the grating diffraction angle θ is given by $\sin \theta = k_g/k$ with $k = \omega/c$ and $k_g = \omega_g/c = 2\pi/d$. The

frequency ω_g is the grating cutoff frequency and d is the spacing between grating fringe. Let us assume (correctly) that we can neglect the short external sections of path length between each grating and its associated mirror compared to the overall cavity path length. Let us also assume (plausibly but erroneously) that we can neglect the small but wavelength-dependent phase shift between the incident and diffracted beams that will occur each time the beam is diffracted through one of the thin gratings, as compared to the large phase delay along the long beam path L . The total round-trip phase delay in the cavity should then be given by

$$\phi(\omega) \approx \frac{2\omega L(\omega)}{c} = \frac{2\omega^2 L_0}{\sqrt{\omega^2 - \omega_g^2}}. \quad (2)$$

This phase shift is plotted as the dashed line in Fig. 1b. If this were indeed the correct dispersion curve for the cavity, the axial mode resonances ω_q , indicated by the points where this curve is intersected by the horizontal lines $\phi(\omega_q) = q2\pi$, would have very interesting dispersion properties.

The error in this analysis lies in neglecting the added grating phase delay at grating G_2 , which though small at any one frequency will in fact have a large cumulative dependence on frequency because of the wavelength dependence of the grating intersection point $a = a(\omega)$. To demonstrate this, suppose, for simplicity, that grating G_1 is a thin transmission grating with amplitude transmission $T_1(x) = T_0 + \Delta T \cos(k_g x + \phi_1)$. The phase ϕ_1 specifies a possible offset of the centermost grating fringe maximum with respect to the x, z coordinate system in Fig. 1a. Sup-

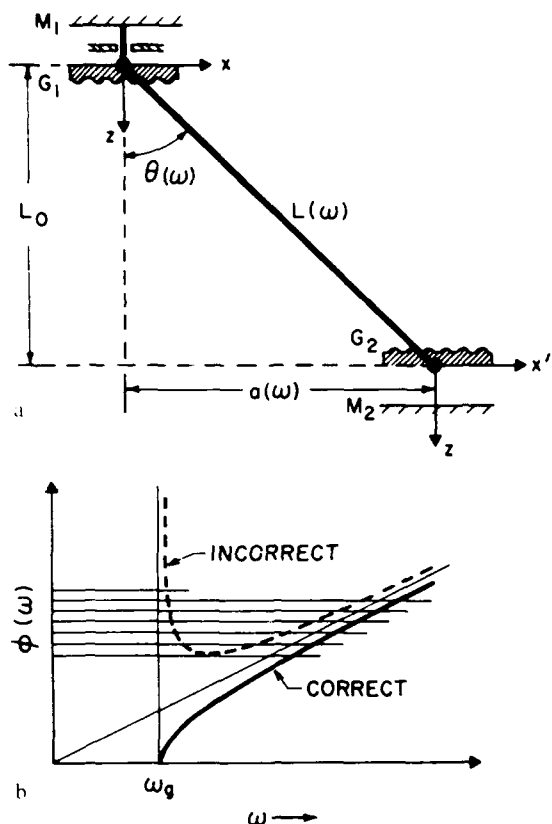


Fig. 1. (a) Analytical model for laser resonator with intracavity grating pair. (b) Correct and incorrect round-trip phase shifts versus frequency

pose we use the general form $u(s, t) = \exp[j(\omega t - ks)]$ for propagation through distance s along the beam path, and assume that the beam path intersects the thin grating G_1 at the origin of the x, z coordinate system. The first-order diffracted wave sent out at angle θ along the path $L(\omega)$ will then pick up a net phase delay of $+\phi_1$ in passing once through this grating, and the same phase delay once again when it comes back through this grating the second time.

We can assume that grating G_2 has a similar amplitude transmission $T_2(x) = T_0 + \Delta T \cos(k_g x + \phi_2)$ in the same coordinate system, with this grating having possibly a different transverse position shift ϕ_2 . To evaluate the phase delay that the beam experiences on passing through this grating, however, we need to write

the transmission in a transverse coordinate x' centered on the beam intersection point with grating G_2 , i.e., $x' = x - a$, where $a = L_0 \tan \theta$. This is given by $T_2(x') = T_0 + \Delta T \cos(k_g x' + k_g a + \phi_2)$. Since the beam is diffracted toward $+x$ at grating G_1 but toward $-x$ at grating G_2 , in the latter case it picks up an added phase delay $-(\phi_2 + k_g a)$ (modulo 2π) on each pass through the grating G_2 . The complete round-trip phase delay, including grating contributions, is then

$$\begin{aligned} \phi &= \frac{2\omega L(\omega)}{c} + 2(\phi_1 - \phi_2 - k_g a) \\ &= \frac{2\sqrt{\omega^2 - \omega_g^2} L_0}{c} + 2(\phi_1 - \phi_2). \end{aligned} \quad (3)$$

This is plotted as the solid curve in Fig. 1b.

When the signal frequency ω is reduced down toward the grating cutoff frequency ω_g , the angle θ approaches 90° , and the beam path length $L(\omega)$ becomes very long. It is then very tempting to conclude that the round-trip phase delay $\phi(\omega)$ must become very large, as indicated by (2) and by the dashed curve in Fig. 1b. The missing point in this analysis is, however, that the phase shift on passing through G_2 also changes rapidly, and with opposite sign, because of the changing intersection point on grating G_2 . Hence, the total round-trip phase delay actually approaches zero, as indicated by (3) and by the solid curve in Fig. 1b, even though the physical path length approaches infinity.

One can confirm from other analytical approaches [3], or simply from consideration of the group delay properties associated with the two curves in Fig. 1b, that the solid curve must be the correct result. This curve then leads to a compression of the axial mode spacings near the grating cutoff frequency ω_g , but not to the widened axial mode spacings or negative-dispersion axial modes we previously predicted. I regret the errors that led to these incorrect conclusions.

References

1. K.A. Stankov: Appl. Phys. B **40**, 103-105 (1986)
2. K.A. Stankov, A.E. Siegman: postdeadline paper, 1986 Conference on Lasers and Electro-Optics (CLEO), San Francisco, CA (June 1986)
3. E.R. Treacy: IEEE J. QE-5, 454-458 (1969)

Advances in Laser Resonator Design

Using Variable Reflectivity Mirrors

A. E. Siegman
Edward L. Ginzton Laboratory
Stanford University
Stanford, California 94305

to appear in
Tutorials in Optics No. 1
presented at the
1987 OSA Annual Meeting

Advances in Laser Resonator Design Using Variable Reflectivity Mirrors

A. E. Siegman
Edward L. Ginzton Laboratory
Stanford University
Stanford, California 94305

Abstract

The best overall design approach at present for obtaining single transverse mode operation in a laser oscillator with large mode volume, good output beam quality, minimum diffraction effects, and arbitrary output coupling is the use of a geometrically unstable laser cavity, combined with a transversely varying reflectivity mirror (VRM) or transversely variable output coupler. Though this approach to laser resonators has only begun to be employed within the past few years, initial demonstrations have been notably successful. In this tutorial we summarize the background analytical methods and the basic design considerations that should be followed to achieve desired laser performance using the most elementary form of variable-reflectivity laser resonator.

I. INTRODUCTION

1.1 Variable-Reflectivity-Mirror (VRM) Optical Resonators

Laser resonators have in the past been divided into two broad categories: stable and unstable. These labels have nothing to do with the fundamental stability of the laser action that can take place in either type of resonator. They refer instead to whether a slightly off-axis optical ray in the laser cavity will oscillate in a stable fashion about the cavity axis, or will diverge from the axis in an exponentially unstable fashion on repeated bounces.

Stable optical resonators display a set of (almost) ideal Hermite-gaussian or Laguerre-gaussian transverse eigenmodes which have very low diffraction losses and are easy to describe analytically. These modes typically have very slender transverse profiles, however, making it difficult to extract all the laser energy from a larger-diameter laser rod or laser tube using only a lowest-order stable eigenmode.

Unstable resonators by contrast have a set of lowest and higher-order transverse modes which are

very complicated in form, with many complex diffraction ripples that are difficult to describe analytically. Unstable resonator eigenmodes can therefore usually be determined only with the aid of more or less complicated numerical calculations. Unstable resonators have in most cases large diffraction output coupling, which can be used as the primary output coupling from the laser device. They also have in general large mode volumes, good transverse mode discrimination, and generally good far-field beam quality. Unstable resonators have therefore been widely employed in larger mode volume laser oscillators which have sufficient round-trip gain to operate effectively with the larger output couplings inherent in the unstable resonator concept.

It has long been known that tapering the reflectivity of an unstable resonator mirror near the mirror edge can improve the mode character and smooth out some of the irregularities in this class of resonators. In fact, the optimum laser cavity design yet proposed appears to be a generalization of this approach, in which one uses geometrically *unstable* cavity optics to produce a round-trip beam magnification and hence large mode volume, together with a smooth, often gaussian transverse variation in mirror reflectivity or in output coupling to limit the transverse magnification and to give good mode discrimination between lowest and higher-order modes. This variable-reflectivity-mirror or VRM approach to laser cavity design can combine all the advantages of stable and unstable optical resonators, producing simultaneously a clean lowest-order gaussian mode inside the laser cavity, good discrimination against higher-order modes, a large transverse mode volume, and operation with either high or low values of output coupling.

Resonators in this category provide a more or less new type of laser cavity design which has only begun to be used in recent years. Because of their useful properties for many types of lasers, this class of VRM resonators appears likely to play an increasingly important role in future laser designs.

1.2 Objectives of this Tutorial

Given this background information, this tutorial has three objectives:

- 1) To review briefly the properties of complex-valued *ABCD* matrices or "complex ray matrices" as simple but powerful tools for describing paraxial optical beams and resonator modes, not just in a geometrical ray picture, but in the most complete and

general wave picture of paraxial wave optics including diffraction.

2) To show how not just two but four general categories of stable and unstable optical resonators arise naturally from the use of these complex-valued $ABCD$ matrices.

3) And, perhaps most important, to provide a set of design procedures for implementing the newest category of geometrically unstable variable-reflectivity-mirror (VRM) laser resonators as described above.

Numerous individuals have contributed to the evolution of the material summarized here; a more or less complete bibliography is given at the end of the chapter. Much of the introductory material in this tutorial is also presented in more detail in the author's textbook on *Lasers*, and occasional references to specific chapters in that book will be given.

II. BASIC TOOLS: THE COMPLEX ABCD MATRIX METHOD

The basic tools for analyzing all possible elementary forms of paraxial (or small-angle) optical resonators are supplied by complex-valued paraxial optical ray matrices or $ABCD$ matrices. Huygens' integral for optical propagation within any such resonator can be written in terms of these $ABCD$ matrices, and complex-valued matrix elements can be used to represent radially tapered gaussian apertures or variable-reflectivity mirrors. A complete set of generalized Hermite-gaussian beam functions with complex-valued scale factors in both the gaussian and Hermite polynomial functions can also always be found as eigensolutions (i.e., transverse eigenmodes) of these resonators. In this section we give a brief but fairly complete review of how these tools are developed.

2.1 Paraxial Ray Matrices

Consider an optical ray which passes through a thin lens of focal length f_1 , then a free-space interval of length L , and then a second lens of focal length f_2 , as shown in Figure 1. If we use x_1 and x'_1 to denote the transverse displacement and slope of this ray with respect to the optical axis (z axis) in this optical system at the transverse plane z_1 , and if the slope x'_1 of the ray is not too large, then these ray coordinates will transform on going through a thin lens in the linear fashion

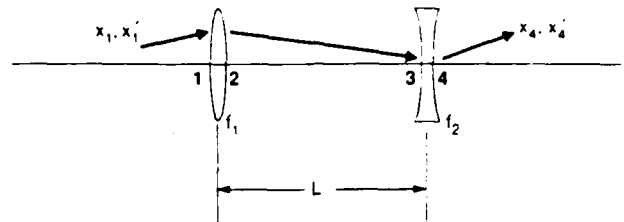


Fig. 1. Example of a paraxial ray matrix system.

$$\begin{aligned} x_2 &= x_1 \\ x'_2 &= x'_1 - (1/f)x_1 \end{aligned} \quad (1)$$

Similarly they will transform on propagating through a free-space length L in the form

$$\begin{aligned} x_3 &= x_2 + Lx'_2 \\ x'_3 &= x'_2 \end{aligned} \quad (2)$$

Suppose we write these transformations in matrix form and then cascade the effects of all the elements in sequence. The total ray transformation going from the input reference plane z_1 just before the first lens to an output reference plane z_4 just after the second lens can then be written in the matrix form

$$\begin{aligned} \begin{bmatrix} x_4 \\ x'_4 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -1/f_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} \\ &= \begin{bmatrix} 1 - L/f_1 & L \\ (L - f_1 - f_2)/f_1 f_2 & 1 - L/f_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} \\ &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} \end{aligned} \quad (3)$$

The overall ray matrix, or so-called $ABCD$ matrix, for the optical system is just the matrix product of the matrices for individual optical elements cascaded together as illustrated above. A thin lens has an $ABCD$ matrix with elements $[1, 0, -1/f, 1]$; a free-space section has elements $[1, L, 0, 1]$; and all other elementary paraxial elements such as curved mirrors, curved dielectric interfaces, graded-index "ducts" or channels have similarly simple $ABCD$ matrices, which can be found in many references [e.g., *Lasers*, Chap. 15]. The total transformation of optical ray coordinates through any cascaded multi-element paraxial optical system can be calculated by elementary matrix multiplication, with the result expressed as a single overall $ABCD$ matrix.

There are a few minor additional complexities to this picture. First of all, inside any dielectric medium with index of refraction n , the slope variable x' should really be interpreted as a "reduced slope" defined by $x'_i \equiv (1/n_i)(dx_i/dz)$ where dx_i/dz is the true ray slope and n_i the index of refraction at the corresponding plane $z = z_i$. With this interpretation, the change in true slope due to Snell's law on passing through a planar dielectric interface is incorporated into the definition of x' , and the ray matrix for a planar dielectric interface is just the identity matrix $[1, 0, 0, 1]$. If we use this convention, all ray matrices then have unity determinant, $AD - BC \equiv 1$.

Second, in an axially symmetric system one can write the same $ABCD$ matrices to apply to either of the transverse coordinates, i.e., one can apply them to either x, x' or y, y' . In an astigmatic paraxial system, however, with principle axes oriented along the x and y directions, one must keep track of separate A_x, B_x, C_x, D_x and A_y, B_y, C_y, D_y matrices governing the x, x' and y, y' transverse directions. There is no cross-coupling between the two transverse directions in this case.

Finally, there even exist "nonorthogonal" (twisted) paraxial systems in which no overall perpendicular pair of principal axes can be found. In these nonorthogonal systems there is an inextricable (but still linear) coupling between all four ray coordinates x, x', y , and y' . Nonplanar ring resonators are one practical example of this useful but slightly more complicated nonorthogonal situation.

And, a final warning: It's easy to forget that the ray matrices must be cascaded in reverse order to the order in which the individual paraxial elements are encountered!

2.2 Huygens' Integral and ABCD Matrices

Equations (1) and (2) look like, and in fact are, just elementary ray optics or geometrical optics. The $ABCD$ formalism nonetheless contains much more than this—in fact, it contains all of diffraction theory or scalar wave optics, at least in the paraxial approximation. Suppose for example that an arbitrary scalar paraxial optical wave of the form $\tilde{u}(x, y, z) \exp[j(\omega t - kz)]$, with an arbitrary input amplitude and phase variation $\tilde{u}_1(x_1, y_1)$, is propagated through any paraxial optical system such as that shown in Figure 2, starting at an input plane $z = z_1$. One can then show that the complex phasor amplitude $\tilde{u}_2(x_2, y_2)$ at the output plane $z = z_2$

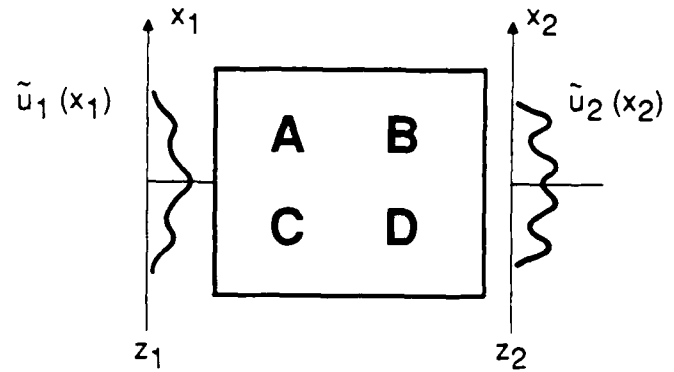


Fig. 2. Huygens' integral through a general paraxial system.

will be given (in the paraxial or Fresnel approximation) by a generalized Huygens' integral written in the form

$$\tilde{u}_2(x_2, y_2) = \int \int \tilde{K}(x_2, x_1) \tilde{K}(y_2, y_1) \tilde{u}_1(x_1, y_1) dx_1 dy_1 \quad (4)$$

where the Huygens' kernel $\tilde{K}(x_2, x_1)$ for either transverse direction can be expressed in terms only of the corresponding overall $ABCD$ matrix of the paraxial system, in the general form

$$\tilde{K}(x_2, x_1) \equiv \sqrt{\frac{j}{B\lambda_0}} \exp \left[-j \frac{\pi}{B\lambda_0} (Ax_1^2 - 2x_1x_2 + Dx_2^2) \right] \quad (5)$$

Thus, in one transverse plane (for simplicity) we have

$$\tilde{u}_2(x_2) = \sqrt{\frac{j}{B\lambda_0}} \int_{-\infty}^{\infty} \exp \left[-j \frac{\pi}{B\lambda_0} (Ax_1^2 - 2x_1x_2 + Dx_2^2) \right] \times \tilde{u}_1(x_1) dx_1 \quad (6)$$

where $ABCD \equiv A_x B_x C_x D_x$. Equations (5) and (6) are evidently a generalization of the usual Huygens' integral formula in which the kernel is written as

$$K(x_2, x_1) = (j/L\lambda)^{1/2} \exp[-j(\pi/L\lambda)(x_2 - x_1)^2] \quad (7)$$

They permit one to propagate a general optical beam $\tilde{u}(x, y, z)$, not just through a free-space section, but

through a complete multi-element paraxial optical system in a single step, using only the overall $ABCD$ matrix of the system, without necessarily knowing the details of the internal optics. The essential approximations in developing this generalized form of the Huygens' integral are:

- The propagating beam, which has the general form $\tilde{u}(x, y, z) \exp[j(\omega t - kz)]$ in the system, must be paraxial. This means in mathematical terms that $|\partial \tilde{u} / \partial z| \ll |k \tilde{u}|$, and in physical terms that the plane-wave components of $\tilde{u}(x, y, z)$ must be limited to angles less than about 30 degrees from the optical axis.
- The optical system must contain only paraxial optical elements (lenses, free space sections, curved mirrors and dielectric interfaces, and quadratic ducts), with no internal hard-edged apertures or sharp transverse discontinuities (though there can be arbitrary aperturing at the input plane z_1).
- The wavelength λ_0 in the equations above should be the free-space (i.e., vacuum) wavelength, even though the input and output planes may be inside dielectric media of index n_1 and n_2 .

Within these limits, Equations (4) through (7) provide extraordinarily useful tools. Multielement diffraction theory is reduced to 2×2 matrix multiplication, followed by a single generalized Huygens' integral. (There is also a slightly more complex 4×4 generalization of this to take care of the more general twisted or nonorthogonal case.)

2.3 Gaussian Optical Beam Propagation

Let us now consider in particular the classic gaussian optical beam, beloved of laser designers and illustrated in Figure 3. Such a beam has a transverse form given in one transverse dimension by

$$\tilde{u}(x) = \exp \left[-j \frac{\pi x^2}{R\lambda} - \frac{x^2}{w^2} \right] = \exp \left[-j \frac{\pi x^2}{\tilde{q}\lambda} \right] \quad (8)$$

The gaussian \tilde{q} parameter appearing in the second exponential is a complex generalization of the radius of curvature R which characterizes a simple uniform spherical wave. In fact, the gaussian \tilde{q} parameter is related to the wavefront radius R and the gaussian spot size w by

$$\frac{1}{\tilde{q}} \equiv \frac{1}{R} - j \frac{\lambda}{\pi w^2} \quad (9)$$

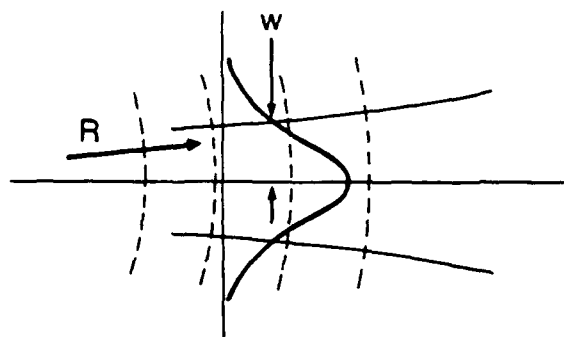


Fig. 3. Spot size w and radius of curvature R for a gaussian optical beam.

In the sign convention used here, the wavefront radius of curvature R is defined so that $R > 0$ corresponds to a wave which diverges in the direction of travel, and the spot size w is the $1/e$ spot size for the field amplitude \tilde{u} , not the intensity $|\tilde{u}|^2$.

Let us now put a gaussian beam $\tilde{u}_1(x_1)$ in the form of Equation (8), with an initial beam parameter $\tilde{q} = \tilde{q}_1$, into the generalized Huygens' integral of Equation (6) and do the integration. The integrand is then a complex-valued quadratic exponent, for which one can use the formula $\int_{-\infty}^{\infty} \exp[-ay^2 - 2by] dy = \sqrt{\pi/a} \exp(b^2/a)$. The output beam $\tilde{u}_2(x_2)$ after passing through the arbitrary $ABCD$ system will turn out to be still a gaussian beam like Equation (6), but with a new output gaussian beam parameter \tilde{q}_2 given by

$$\tilde{q}_2 = \frac{A\tilde{q}_1 + B}{C\tilde{q}_1 + D} \quad (10)$$

In other words, for gaussian beam propagation through an arbitrary multi-stage paraxial system one does not even have to do the Huygens' integration; one can just use Equation (10).

Again, a few small complications: If the input and output planes happen to be inside dielectric media, then the wavelength λ appearing in Eqs. (8) and (9) refers to the wavelength in the medium, i.e., λ may be different at each end. If this is the case then Equation (10) must be written more generally as

$$\tilde{q}_2/n_2 = \frac{A(\tilde{q}_1/n_1) + B}{C(\tilde{q}_1/n_1) + D} \quad (11)$$

For these situations it might be better to define a "reduced \tilde{q} parameter," call it \hat{q} , which is given at

each plane by

$$\frac{1}{\hat{q}_i} \equiv \frac{n_i}{\hat{q}_i} \equiv \frac{n_i}{R_i} - j \frac{n_i \lambda_i}{\pi w_i^2} \quad (12)$$

where n_i , R_i , w_i and λ_i are all the real values in the dielectric medium at that plane, and $n_1 \lambda_1 = n_2 \lambda_2 = \lambda_0$. In terms of this \hat{q} -hat parameter, Equation (10) would then become, universally,

$$\hat{q}_2 = \frac{A\hat{q}_1 + B}{C\hat{q}_1 + D} \quad (13)$$

The bad news is that the less general definition of Equation (9) is firmly entrenched compared to the preferable definition of Equation (12). The good news is, in real problems we almost always go from one plane in free space to another plane in free space, so the distinctions involved in Equations (9) through (13) don't matter that much anyway.

2.4 Complex-Valued ABCD Matrices

We can next develop a complex-valued (instead of purely real-valued) generalization of all the above results. As a starting example, we can note that a conventional curved mirror with radius of curvature R_c , or equivalently a thin lens with focal length $f \equiv R_c/2$, will impose an added transversely varying phase shift on a wave bouncing off it which is given by

$$\begin{aligned} \tilde{u}_2(x) &= \exp \left[+j \frac{2\pi x^2}{R_c \lambda} \right] \times \tilde{u}_1(x) \\ &= \exp \left[+j \frac{\pi x^2}{f \lambda} \right] \times \tilde{u}_1(x) \end{aligned} \quad (14)$$

Such a mirror or lens is represented by a purely real $ABCD$ matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2/R_c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \quad (15)$$

Note: In our sign conventions $f > 0$ means a converging lens, and hence $R_c > 0$ means a mirror concave toward the arriving optical beam; the sign conventions for mirror radii and wavefront radii are thus slightly different.

Consider now a gaussian variable reflectivity mirror (VRM) by which we mean a mirror having a transversely varying amplitude reflectivity $\tilde{\rho}(x)$ given by

$$\tilde{u}_2(x) = \tilde{\rho}(x) \times \tilde{u}_1(x) = \exp \left[-\frac{x^2}{w_{ga}^2} \right] \times \tilde{u}_1(x) \quad (16)$$

where w_{ga} is the $1/e$ radius for the amplitude reflection variation across the mirror. Alternatively, $\tilde{\rho}(x)$ could equally well be the amplitude transmission through a gaussian aperture with transversely varying transmission. If we compare Equations (14) and (16) we see that such a gaussian VRM or gaussian aperture acts on a wave just like a curved mirror or thin lens having an imaginary-valued focal length f' given by

$$\frac{1}{f'} \equiv \frac{2}{R'_c} \equiv j \frac{\lambda}{\pi w_{ga}^2} \quad (17)$$

Hence, we are led to propose that the gaussian VRM or aperture of Equation (16) should be characterized by a complex-valued $ABCD$ matrix given, in analogy to Equation (15), by

$$\begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -j\lambda/\pi w_{ga}^2 & 1 \end{bmatrix} \quad (18)$$

More careful examination then shows that, in fact:

- This is a very useful way to describe gaussian variable-reflectivity mirrors or variable-transmission apertures in the $ABCD$ formalism (even if it is not at all clear what a complex-valued geometrical ray might look like).
- One can cascade-multiply these complex-valued $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrices with each other or with other real $ABCD$ matrices in exactly the same fashion as described above.
- And, in fact, all of the Huygens' integral and gaussian-beam formulas given in this tutorial thus far will remain equally true for such complex-valued $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrices in paraxial systems containing gaussian mirrors or apertures.

The end result is that this complex-valued ray matrix approach gives the most general possible form of paraxial wave optics, with results which are valid for any system in which the transmission or reflection functions for all the individual elements may have arbitrary complex-valued quadratic variations in their exponents.

2.5 Complex Hermite-Gaussian Modes

We can now look also for a more general complex-valued set of lowest and higher-order Hermite-gaussian eigenfunctions, to replace (or rather to extend) the conventional Hermite-gaussian functions of stable resonator theory. That is, let us propose that the most general family of lowest- and higher-order

Hermite-gaussian modes $\tilde{u}_n(x; \tilde{q}, \tilde{v})$ can be written, in one transverse coordinate, in the form

$$\tilde{u}_n(x; \tilde{q}, \tilde{v}) \equiv \tilde{\alpha}_n \tilde{v}^n H_n \left(\frac{\sqrt{2}x}{\tilde{v}} \right) \exp \left(-j \frac{\pi x^2}{\tilde{q}\tilde{\lambda}} \right) \quad (19)$$

where all the parameters $\tilde{\alpha}$, \tilde{v} and \tilde{q} can now in general be complex-valued quantities. The parameter \tilde{q} in this case still retains exactly the same interpretation in terms of beam radius R and spot size w as in Equation (9). The parameter \tilde{v} , however, is a new and potentially complex-valued scale factor appearing in the argument of the Hermite-polynomials $H_n(\sqrt{2}x/\tilde{v})$. The family of functions in Equation (19) provides a completely generalized form of the more conventional Hermite-gaussian functions $H_n(\sqrt{2}x/w) \exp(-x^2/w^2)$ that are well known, for example, as the mode functions in conventional stable optical resonators, or as the quantum wave functions for a quantized simple harmonic oscillator.

The family of functions in Equation (19) forms a mathematically complete basis set: these functions can be used to expand any arbitrary complex function $\tilde{f}(x)$ in the form $\tilde{f}(x) = \sum_n \tilde{\alpha}_n \tilde{u}_n(x; \tilde{q}, \tilde{v})$, for any arbitrary choice of \tilde{q} and \tilde{v} , although the "best" choice for these parameters may depend on the physical problem, as we will see below.

Note that the lowest-order symmetric and anti-symmetric modes in this family, that is, the modes \tilde{u}_n for $n = 0$ and $n = 1$, do not in fact depend on the value of the complex scale factor \tilde{v} . In the higher-order modes, however, for $n \geq 2$, the use of a complex rather than purely real value for \tilde{v} will lead to some interesting results. The higher-order modes, for example, will no longer be purely spherical waves, and will no longer have sharp nulls in the Hermite functions. Instead the phase front will be "wrinkled," and the usual zeros in the Hermite polynomials will be partly filled in, as a result of the complex polynomial variation of $H_n(\sqrt{2}x/\tilde{v})$ with x . Fortunately, it is seldom necessary to actually use these higher-order modes in practical situations.

One important fundamental property of the complex Hermite-gaussian functions given in Equation (19) is the following. Suppose we send an optical beam having a transverse profile $\tilde{E}_1(x_1)$ given by any single one of these complex Hermite-gaussian modes, e.g. $\tilde{E}_1(x) = \tilde{\alpha}_{1n} \tilde{u}_n(x; \tilde{q}_1, \tilde{v}_1)$, into an arbitrary complex-valued $ABCD$ system—that is, into an arbitrary cascade of lens, mirrors, gaussian apertures,

etc. We can then evaluate the beam at the output from this system by evaluating the Huygens' integral of Equation (6), using the overall complex $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ values and the function $\tilde{E}_1(x)$ as input. [Equation (6) remains valid even for complex $ABCD$ matrix elements.] The result turns out to be that the output beam is still a single complex-valued Hermite-gaussian function, $\tilde{E}_2(x) = \tilde{\alpha}_{2n} \tilde{u}_n(x; \tilde{q}_2, \tilde{v}_2)$, with the same order n . The three complex-valued parameters $\tilde{\alpha}_{1n}$, \tilde{q}_1 , \tilde{v}_1 , at the input plane will be converted, however, into three new complex values $\tilde{\alpha}_{2n}$, \tilde{q}_2 , \tilde{v}_2 at the output plane according to the rules

$$\begin{aligned} \tilde{q}_2 &= \frac{\tilde{A}\tilde{q}_1 + \tilde{B}}{\tilde{C}\tilde{q}_1 + \tilde{D}} \\ \tilde{v}_2^2 &= (\tilde{A} + \tilde{B}/\tilde{q}_1)^2 \tilde{v}_1^2 + j \frac{4\tilde{B}}{k_1} (\tilde{A} + \tilde{B}/\tilde{q}_1) \\ \frac{\tilde{\alpha}_{2n}}{\tilde{\alpha}_{1n}} &= \left(\frac{1}{\tilde{A} + \tilde{B}/\tilde{q}_1} \right)^{n+1/2} \end{aligned} \quad (20)$$

The first of these rules we already knew; the second and third are new.

Equations (14) through (20) provide a complete complex-valued generalization of paraxial optical propagation theory. Any arbitrary input function $\tilde{E}_1(x)$ at an input plane $z = z_1$ can be expanded as a sum over n of these functions, i.e. $\tilde{E}_1(x) = \sum_n \tilde{\alpha}_{1n} \tilde{u}_n(x; \tilde{q}_1, \tilde{v}_1)$, using arbitrary choices of \tilde{q}_1 and \tilde{v}_1 . The individual functions $\tilde{\alpha}_{1n} \tilde{u}_n(x; \tilde{q}_1, \tilde{v}_1)$ can each be propagated through the $ABCD$ system to give $\tilde{\alpha}_{2n} \tilde{u}_n(x; \tilde{q}_2, \tilde{v}_2)$, using Eqs. (20), and the output beam is then $\tilde{E}_2(x) = \sum_n \tilde{\alpha}_{2n} \tilde{u}_n(x; \tilde{q}_2, \tilde{v}_2)$. Equations (20) are thus adequate to propagate any arbitrary paraxial optical beam $\tilde{E}(x)$ through any complex-valued paraxial optical system, using only the overall complex $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix of the system.

III. GENERALIZED PARAXIAL RESONATOR THEORY

We can now introduce a quite simple and yet very general and useful analysis of the transverse eigenmodes in the most general possible set of *generalized paraxial optical resonators*, using the tools from the preceding section, as follows.

3.1 Self-Consistent Gaussian Mode Solutions

Suppose we have an arbitrary complex-valued laser resonator of either ring or standing-wave type, and we want to find the transverse eigenmodes of this resonator. To do this, we first pick a reference plane

at any arbitrary but convenient location within the resonator. Just before the output mirror or coupler of the laser is often a convenient point, since we can then evaluate the beam profile just before it comes out of the laser cavity. We can then write down and multiply together all the (possibly complex-valued) $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ elements for one complete round trip within the laser cavity, starting at and returning to this reference plane, to get the complete round-trip $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix for the cavity.

Now, if a certain beam profile $\tilde{u}(x)$ is to be a transverse eigenmode of this laser cavity at this plane (considering only one transverse coordinate for simplicity), then this beam profile should return to this plane after one complete round trip around the cavity with the same shape, i.e., the same complex beam profile as it started out with, although the complex amplitude of the beam may be reduced by a complex constant of magnitude less than unity to account for diffraction losses and phase shifts during the round trip. But, the complex Hermite-gaussian functions $\tilde{u}_n(x; \tilde{q}, \tilde{v})$ of Equation (19) will satisfy this round-trip self-consistency condition, and hence can be the transverse eigenmodes of the laser cavity, if the complex parameters \tilde{q} and \tilde{v} , as related by Equations (20), remain unchanged in one round trip.

In fact, for the two lowest-order modes $n = 0$ and $n = 1$, only the \tilde{q} parameter needs to remain unchanged in one round trip. Hence the round-trip self-consistency condition for the $n = 0$ and $n = 1$ modes reduces to the requirement that after one round trip $\tilde{q}_2 = \tilde{q}_1$, or

$$\tilde{q}_2 = \frac{\tilde{A}\tilde{q}_1 + \tilde{B}}{\tilde{C}\tilde{q}_1 + \tilde{D}} = \tilde{q}_1 \quad (21)$$

Rewriting this in inverted form gives the quadratic equation

$$\left(\frac{1}{\tilde{q}_1}\right)^2 + \frac{\tilde{A} - \tilde{D}}{\tilde{B}} \left(\frac{1}{\tilde{q}_1}\right) + \left(\frac{1 - \tilde{A}\tilde{D}}{\tilde{B}^2}\right) = 0 \quad (22)$$

where we have used the general relationship that $\tilde{A}\tilde{D} - \tilde{B}\tilde{C} = 1$. This equation has two roots given by

$$\frac{1}{\tilde{q}_a}, \frac{1}{\tilde{q}_b} = \frac{\tilde{D} - \tilde{A}}{2\tilde{B}} \mp \frac{1}{\tilde{B}} \sqrt{\left(\frac{\tilde{A} + \tilde{D}}{2}\right)^2 - 1} \quad (23)$$

and either one of these two roots is a potential self-consistent "eigen- \tilde{q} -value" for the optical resonator.

3.2 Confined Gaussian Modes

For the gaussian eigensolutions of Equation (23) to be physically meaningful, however, the resulting \tilde{q} value should describe a radially decreasing gaussian mode in the form of Equation (9), i.e., the imaginary part of $1/\tilde{q}$ must be negative in order to give a truly negative value of the imaginary portion $-j\lambda/\pi w^2$. Only one of the two subscripts a, b or one of the choices of the \mp sign in Equation (23) will normally give such a physically meaningful, radially decreasing solution, which we will refer to as a transversely confined eigensolution.

We will see what this implies for some real resonator cases shortly. Let us push on with a little additional math for a few more paragraphs, however, and then get to these real cases.

3.3 Gaussian Mode Eigenvalues

Suppose next that we launch one of the Hermite-gaussian eigenmodes given by Equation (23) inside the resonator with the appropriate confined and self-consistent value of \tilde{q}_a or \tilde{q}_b given by (23). This gaussian beam pattern will return after one round trip with the same shape, or the same beam profile, but with its amplitude reduced by the constant factor

$$\begin{aligned} \frac{\tilde{\alpha}_{n2}}{\tilde{\alpha}_{n1}} &= \left[\frac{1}{\tilde{A} + \tilde{B}/\tilde{q}_{a,b}} \right]^{n+1/2} = \left[\frac{1}{\tilde{m} \mp \sqrt{\tilde{m}^2 - 1}} \right]^{n+1/2} \\ &= [\tilde{m} \pm \sqrt{\tilde{m}^2 - 1}]^{n+1/2} \equiv [\tilde{\lambda}_a, \tilde{\lambda}_b]^{n+1/2} \end{aligned} \quad (24)$$

Hence the complex round-trip amplitude transmission for each mode is completely determined by a single important parameter of the resonator, namely the half-trace parameter \tilde{m} defined by

$$\tilde{m} \equiv \frac{\tilde{A} + \tilde{D}}{2} \quad (25)$$

Note that in Equation (24) the ratios $\tilde{\lambda}_a$ and $\tilde{\lambda}_b$ refer to the choice of upper or lower signs in the formulas, just as \tilde{q}_a and \tilde{q}_b refer to the upper and lower signs in Equation (23). The complex-valued ratios $\tilde{\lambda}_a$ and $\tilde{\lambda}_b$ defined by Equation (24) are referred to as the matrix eigenvalues or perturbation eigenvalues of the resonator. Note also that \tilde{m} is potentially complex, if \tilde{A} and \tilde{D} have complex values.

The important result in Equation (24) is that the amplitude of the n -th order self-consistent gaussian mode characterized by beam parameter \tilde{q}_a or \tilde{q}_b

will be multiplied on each round trip by the complex eigenvalue $\tilde{\lambda}_a$ or $\tilde{\lambda}_b$ raised to the $n + 1/2$ power (assuming the calculation is done in one transverse coordinate only).

3.4 Perturbation Stability

There is an additional and perhaps more subtle requirement on these gaussian eigensolutions and eigenvalues. The eigenmodes of a real resonator should also be perturbation-stable against small initial perturbations in the mode profile. That is, suppose we launch into the resonator a complex-valued Hermite-gaussian mode whose initial \tilde{q} value is slightly perturbed from the confined and self-consistent value \tilde{q}_a or \tilde{q}_b , so that $\tilde{q}_1 = \tilde{q}_{a,b} + \Delta\tilde{q}_1$; and then we use Equation (20) to evaluate the output \tilde{q} value $\tilde{q}_2 = \tilde{q}_{a,b} + \Delta\tilde{q}_2$ after one round trip. We then obtain, from Equation (20),

$$\tilde{q}_2 = \frac{\tilde{A}(\tilde{q}_{a,b} + \Delta\tilde{q}_1) + \tilde{B}}{\tilde{C}(\tilde{q}_{a,b} + \Delta\tilde{q}_1) + \tilde{D}} = \tilde{q}_{a,b} + \Delta\tilde{q}_2 \quad (26)$$

With a small amount of algebra we can find that, for $|\Delta\tilde{q}| \ll |\tilde{q}_{a,b}|$, this leads to the result

$$\frac{\Delta\tilde{q}_2}{\Delta\tilde{q}_1} \approx [\tilde{\lambda}_a, \tilde{\lambda}_b]^2 \quad \text{for} \quad \tilde{q}_1 = \tilde{q}_{a,b} + \Delta\tilde{q}_1 \quad (27)$$

(The notation $[\tilde{\lambda}_a, \tilde{\lambda}_b]$ means $\tilde{\lambda}_a$ applies for perturbations about \tilde{q}_a and $\tilde{\lambda}_b$ for perturbations about \tilde{q}_b .) If the resonator eigenmode is to be physically acceptable, not only should its gaussian parameter \tilde{q}_a or \tilde{q}_b be self-consistent after each round trip, and not only should its wave amplitude $\tilde{\alpha}_n$ decrease by a ratio $\tilde{\lambda}_a^{n+1/2}$ or $\tilde{\lambda}_b^{n+1/2}$ which should be less than unity on each round trip, but also any small perturbation $\Delta\tilde{q}$ to that mode's gaussian parameter should also decrease by a ratio $\tilde{\lambda}_a^2$ or $\tilde{\lambda}_b^2$ that is less than unity in magnitude on each round trip.

From examination of Equation (24)—or more fundamentally, from the fact that even the complex-valued $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix always has unity determinant—we can see that the two eigenvalues $\tilde{\lambda}_a$ and $\tilde{\lambda}_b$ of any arbitrary resonator always satisfy the condition that

$$\tilde{\lambda}_a \times \tilde{\lambda}_b = \left[\tilde{m} + \sqrt{\tilde{m}^2 - 1} \right] \times \left[\tilde{m} - \sqrt{\tilde{m}^2 - 1} \right] \equiv 1 \quad (28)$$

Hence either one of the two eigensolutions \tilde{q}_a or \tilde{q}_b is stable, with the corresponding $|\tilde{\lambda}| < 1$, and the other

is not (since it has $|\tilde{\lambda}| > 1$), or else both eigenvalues are marginally stable with $|\tilde{\lambda}_{a,b}| = 1$.

3.5 Generalized Paraxial Resonator Categories

The essential requirements for a generalized paraxial optical resonator to have at least the two lowest-order $n = 0$ and $n = 1$ complex-valued Hermite-gaussian modes as physically realistic and meaningful transverse eigenmodes thus reduce to:

- 1) The gaussian \tilde{q} parameter for these modes must be self consistent after each round trip. Hence, it can be either one of the two solutions \tilde{q}_a or \tilde{q}_b given by Equation (23), where all of the quantities in that expression are potentially complex-valued.
- 2) But, such a mode can only be physically realistic if the imaginary part of either $1/\tilde{q}_a$ or $1/\tilde{q}_b$ is negative, corresponding to a negative value for the quantity $-1/w^2$ in Equation (9). Hence, the mode must be a confined rather than a transversely diverging gaussian beam.
- 3) And, finally, the eigenvalue $\tilde{\lambda}_a$ or $\tilde{\lambda}_b$ corresponding to this confined gaussian mode, as given by the expression

$$\tilde{\lambda}_{a,b} = \tilde{m} \pm \sqrt{\tilde{m}^2 - 1} \quad (29)$$

must have a magnitude $|\tilde{\lambda}| \leq 1$, so that the confined mode is also (at least marginally) perturbation-stable.

Examination of these expressions then shows that all paraxial optical resonators can be conveniently separated into not just two, but four, distinct categories, as follows:

1. Real, geometrically stable resonators ($ABCD$ purely real, $m^2 < 1$)

The most straightforward gaussian-beam solutions arise when the round-trip $ABCD$ matrix is purely real, meaning that there are no gaussian apertures or variable reflectivity mirrors, and the matrix half-trace $m \equiv (A + D)/2$ has magnitude less than unity. The two eigen- \tilde{q} -values then have the form

$$\frac{1}{\tilde{q}_a}, \frac{1}{\tilde{q}_b} = \frac{D - A}{2B} \mp j \frac{\sqrt{1 - m^2}}{B} \quad (30)$$

and the eigenvalues have the form

$$\tilde{\lambda}_a, \tilde{\lambda}_b = m \pm j \sqrt{1 - m^2} = \cos \theta \pm j \sin \theta = e^{\pm j\theta} \quad (31)$$

where θ is an angle defined by $\theta = \cos^{-1} m$. We can see that, depending on the sign of B , one or the other of \tilde{q}_a or \tilde{q}_b must obviously represent a confined gaussian beam. Moreover, the eigenvalue corresponding to that \tilde{q} value will be at least marginally stable, i.e., perturbations to this mode will oscillate but not grow on successive round trips.

The confined and perturbation-stable (or at least not perturbation-unstable) solutions found in these cases are just the conventional and well-known real Hermite-gaussian modes in ordinary stable laser cavities. The $ABCD$ approach enables us to calculate these Hermite-gaussian modes easily, not only for the usual two-mirror cavity case with $0 \leq g_1 g_2 \leq 1$, but for any complex multi-element stable cavity with multiple mirrors, internal lenses and telescopes, and so forth. Pick a reference plane; evaluate the round-trip $ABCD$ matrix starting from that plane; evaluate Equations (30) and (31); pick the confined \tilde{q} value; and you have the gaussian beam solution for that resonator. All the higher-order modes with $n > 1$ will also exist in this case, with $\tilde{v} \equiv w$, i.e., with a purely real scale factor.

Within the limits of this analysis, these Hermite-gaussian modes all appear to be totally lossless, i.e., they have $|\lambda| \equiv 1$. This occurs because the ideal paraxial analysis implicitly assumes that all mirrors have infinite transverse width or diameter. If a finite-diameter mirror or aperture is added to the cavity—for example at the reference plane—these idealized Hermite-gaussian modes will be slightly perturbed, especially near the cavity edges, and the eigenmodes will acquire small but finite diffraction losses. These diffraction losses will increase with decreasing resonator Fresnel number (i.e., as the aperture is made smaller), and also will increase with increasing mode index n .

The exact mode patterns and mode losses with finite-diameter mirrors cannot be found from the simple paraxial mode analysis, but only from a Fox-and-Li-type numerical analysis. This analysis can be set up, however, using the $ABCD$ formalism to write the diffraction integral like Equation (4), going once around the cavity starting from a reference plane at the finite mirror or aperture.

2. Real, geometrically unstable resonators ($ABCD$ purely real, $m^2 > 1$):

A second basic class of resonators occurs when the $ABCD$ elements are again purely real but the

half-trace has value $|m| > 1$. In this case the eigen- \tilde{q} -values have the form

$$\frac{1}{\tilde{q}_a}, \frac{1}{\tilde{q}_b} = \frac{1}{B} \left[\frac{A-D}{2} \mp \sqrt{m^2 - 1} \right] = \frac{1}{R_a}, \frac{1}{R_b} \quad (32)$$

The eigenwaves in this case are purely real spherical waves with no gaussian fall-off, and the corresponding eigenvalues are also purely real numbers given by

$$\tilde{\lambda}_a, \tilde{\lambda}_b = m \pm \sqrt{m^2 - 1} = M \quad \text{and} \quad 1/M \quad (33)$$

The so-called *geometric magnification* M that appears here has magnitude $|M| > 1$, and may correspond to either the eigenvalue $\tilde{\lambda}_a$ or $\tilde{\lambda}_b$, depending on whether $m > +1$ or $m < -1$.

Both "eigenmodes" in this case are unconfined, i.e., both are unbounded uniform spherical waves with radii of curvature R_a and R_b . Further examination shows that one of these waves, which might be called the "diverging-wave" solution, is a uniform spherical wave whose transverse width expands or magnifies transversely by the magnification factor M on each round trip, while its radius of curvature remains unchanged. This diverging-wave solution is in fact also the perturbation-stable solution to the resonator problem, i.e., its corresponding eigenvalue is $\lambda = 1/M$, not $\lambda = M$, so that any perturbations to this solution decrease in magnitude on successive round trips even though the beam width blows up in size on successive round trips. The other "converging-wave" solution demagnifies in transverse width by $1/M$ on each round trip, but is perturbation-unstable, with eigenvalue $\lambda = M$.

These uniform spherical wave solutions are obviously not physically acceptable resonator eigenmodes by themselves, since they have no transverse fall-off in amplitude, and hence are not confined. The diverging spherical wave solutions do, however, form the starting point, or the zero-order approximation, for the more complicated modes that occur in real finite-diameter unstable optical resonators. Suppose one adds a finite-diameter mirror or coupling aperture to a real, geometrically unstable paraxial resonator and evaluates the real transverse eigenmodes, using for example a numerical Fox-and-Li approach. One will find that the phasefronts of the real eigenmodes (the lowest-order and higher-order eigenmodes) look very much like the diverging spherical wave solution R_a or R_b , except that the wavefronts of the real modes will be slightly wrinkled in phase.

The transverse amplitude patterns of the real eigenmodes in a finite-diameter unstable resonator will be very complicated patterns with many Fresnel diffraction ripples. These eigenmode patterns, however, will each expand by more or less the geometrical magnification M on each round trip; be cut back down in size by passing through the finite mirror or aperture; and then expand by M again on the next round trip. The resulting steady-state eigenmode amplitude patterns will be far from uniform, displaying instead a complex pattern of diffraction ripples produced by the strong edge-diffraction effects at the aperture edges.

The real eigenmodes in a finite-diameter unstable resonator are thus based primarily on the perturbation-stable diverging-wave solutions, rather than the perturbation-unstable converging-wave solutions. The converging-wave solutions play a certain secondary role, however, as follows. When the real, magnifying eigenmode in an unstable resonator strikes the finite mirror edge or output aperture, some of the diverging-wave energy gets scattered by edge diffraction so as to produce a small excitation of the converging-wave term. The energy scattered into the converging-wave-like term then converges down near the axis, spreads out again into diverging-wave form due to diffraction (or, to perturbation-instability), and eventually interferes with the diverging-wave part. This interference between converging and diverging terms due to scattering effects at the mirror edges is part of the overall edge diffraction process in the unstable resonator, and accounts for the well-known periodic variation in unstable resonator properties as the resonator Fresnel number is increased at fixed geometric magnification.

3. Complex, confined, perturbation-stable resonators:

Consider next the very important case where the $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix becomes complex-valued, because of the presence of a gaussian mirror or aperture in the resonator. For this particular case to occur, the end mirror must have maximum reflectivity on axis, with increased loss or reduced gain off axis. (A gaussian aperture inside the cavity, with maximum transmission on axis will of course accomplish the same result). One will then find that the self-consistent gaussian parameters \tilde{q}_a or \tilde{q}_b for this cavity are both complex quantities, with one or the other of them having a negative imaginary part and thus corre-

sponding to a confined gaussian mode. Furthermore, the eigenvalue $\tilde{\lambda}_a$ or $\tilde{\lambda}_b$ corresponding to that confined gaussian mode will have magnitude less than unity, i.e., this confined mode will be simultaneously perturbation-stable. (The other mode will then necessarily be both unconfined or transversely divergent, and also perturbation-unstable.)

The resulting class of complex, confined, perturbation-stable resonators is the prime target of this tutorial, since resonators in this class appear to have near-ideal properties for practical laser applications. The optics inside a resonator of this type can be made geometrically unstable or magnifying, like the purely real unstable resonator described above, and this can lead to a large mode volume for the resonator. Yet, the presence of a gaussian VRM or aperture ultimately constrains the mode size, and also causes the mode shape to be an ideal (though potentially large-diameter) gaussian or Hermite-gaussian mode—a mode shape which is particularly “clean” and can be easily treated analytically. The loss discrimination between lowest and higher-order modes in this class of resonators can also be very good. We will examine the properties of these resonators in more detail in the following section.

4. Complex, confined, perturbation-unstable resonators:

Finally, there is the contrasting case where the $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ values are again complex, but the losses in the system have maximum value on axis, decreasing as one goes off axis. This could mean a variable-reflectivity mirror with a minimum reflectivity on axis, or, as another example, a gain tube with a gain that increases as one goes outward from the tube axis. In any case, this aperture acts like Equations (16) and (18) with the minus sign in front of the x^2/w^2 term replaced by a plus sign.

This case again turns out to have one confined and one unconfined \tilde{q} value, just as in the preceding case. However, the confined gaussian mode is now always perturbation-unstable, and the perturbation-stable mode is now always unconfined. These gaussian modes are thus not useful for laser action—even if the confined gaussian mode somehow gets started oscillating, it will always “blow up” in size in response to any small perturbation, as a result of its inherent perturbation-instability.

3.6 Higher-Order Modes: Stability of the \tilde{v} Parameter

We have up to this point avoided much discussion of the \tilde{v} parameter appearing in the higher-order Hermite-gaussian modes with $n \geq 2$, or the value this parameter must have in a stable and self-consistent resonator. In fact, we are most often interested only in the lowest-order $n = 0$ and perhaps the $n = 1$ modes of a resonator, and hence we need not worry much about the \tilde{v} parameter. We can summarize the basic analytical results for this parameter quite briefly, however, as follows.

If a cavity is to have significant and useful higher-order eigenmodes, then the \tilde{v} parameter for these modes must also be self-consistent after each round trip, just like the \tilde{q} parameter. Such self-consistent values, call them \tilde{v}_a and \tilde{v}_b , can in fact be found by setting $\tilde{v}_2 = \tilde{v}_1$ in Equation (20). The confinement criterion is not really relevant for these self-consistent \tilde{v} parameters, since the radial growth or decay of the wave function at large transverse distances is totally controlled by the imaginary part of the $1/\tilde{q}$ factor in the gaussian exponent. One must, however, test the perturbation stability of the \tilde{v}_a and \tilde{v}_b parameters in the same fashion as the perturbation stability of the \tilde{q}_a and \tilde{q}_b parameters.

If one does this, using Equation (20) and the analogs of Equations (26) and (27), one finds the following. Whereas perturbations to \tilde{q}_a grow (or decay) as $\tilde{\lambda}_a^2$, perturbations to \tilde{v}_a grow (or decay) as $\tilde{\lambda}_b^2 \equiv \tilde{\lambda}_a^{-2}$, and the same results are obtained with a and b interchanged for perturbations to \tilde{q}_b and \tilde{v}_b . In other words, either \tilde{q} and \tilde{v} are both marginally stable, with $|\tilde{\lambda}| \equiv 1$ (this is the real and geometrically stable case); or else if \tilde{q} is perturbation-stable ($|\tilde{\lambda}| < 1$), then the matching \tilde{v} is necessarily not perturbation-stable.

In short, these higher-order modes are not of much use in real complex-valued optical resonators.

IV. DESIGNING GAUSSIAN VRM RESONATORS

In this section we come finally to the primary objective of this tutorial: analysis and design procedures for practical laser resonators using gaussian variable reflectivity mirrors. These resonators appear to be so useful that one can predict that in the near future, as practical forms of variable-reflectivity mirrors evolve and become more easily available, they will become the optimum form of cavity design for all types of lasers except (a) the very lowest-gain and

slender-bore types of lasers, where conventional stable resonators are perfectly adequate, and (b) the very highest-power CW lasers which can function only with all-reflective water-cooled optics.

4.1 A Brief Historical Review

It was realized quite early (see References) that tapering the reflectivity of a conventional unstable resonator mirror near the mirror edge might improve the mode character of such resonators. Edge-wave diffraction in a conventional hard-edged unstable resonator is the physical effect responsible for the complex and irregular near-field mode patterns in ordinary unstable resonators—mode patterns which are unaesthetic, even if not really otherwise so objectionable. It is also responsible for the complicated periodic interweaving of the unstable resonator eigenvalues with varying Fresnel number, an effect which continues all the way to $N_F = \infty$ for circular-aperture unstable resonators. Tapering the mirror reflectivity, either in phase or in amplitude, over a range of one Fresnel zone or so near the mirror edge “fuzzes out” these Fresnel edge diffraction effects and thus reduces both of the above effects.

Various forms of tapered, stepped, and apodized mirror designs have been considered for this purpose, a few of these being mentioned in the References. In particular, Zucker, Chester, Suematsu, and Russian authors gave early consideration to VRM resonators with gaussian reflectivity tapers. It was not until the development by Casperson, Yariv, and their coauthors of the generalized complex paraxial analysis summarized in this tutorial, however, that a really flexible and convenient design analysis could be developed and the advantages of these resonators fully appreciated.

It is probably the case that the optimum overall resonator design, especially in terms of internal as well as external mode profiles, can be obtained by the use of more sophisticated VRM reflectivity profiles than the elementary gaussian profile discussed in this tutorial, such as for example, the “hypergaussian” profiles mentioned briefly further on. The simple gaussian reflectivity profile is both so useful and so easy to analyze, however, that we will make it the primary focus of this tutorial.

4.2 Basic Analysis

The simplest and most idealized and yet also quite practical form of reflectivity tapering is the case in

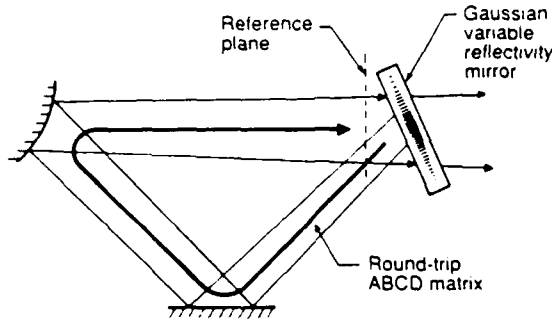


Fig. 4. Analytical model for round-trip propagation in a VRM resonator.

which the output mirror's amplitude or E -field reflectivity $\tilde{\rho}(r)$ has the form

$$\tilde{\rho}(r) = R_0^{1/2} \exp[-r^2/w_{ga}^2] \quad (34)$$

and hence the power reflectivity $R(r)$ from the mirror has the transverse variation

$$R(r) = R_0 \exp[-2r^2/w_{ga}^2] \quad (35)$$

Here R_0 is the power reflectivity and $T_0 \equiv 1 - R_0$ the power transmission at the center of the output mirror. (Practical methods for obtaining a reflectivity variation similar to this are discussed in a later section.)

The basic analysis of a gaussian VRM laser cavity is then very simple. We select a reference plane within the cavity—most conveniently at the plane just before the beam strikes the output mirror, as shown in Figure 4. The solution we obtain will then give us the beam profile just before it exits the cavity. The variable reflectivity mirror itself will be represented by an $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix with elements $[1, 0, -j\lambda/\pi w_{ga}^2, 1]$ as in Equation (18), and we can represent the remainder of the complete round trip from the reference plane around the cavity and back again by a purely real $ABCD$ matrix. The complete round-trip around the cavity, including the gaussian aperture, can then be represented by the total complex-valued $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ matrix

$$\begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ -j\lambda/\pi w_{ga}^2 & 1 \end{bmatrix} \quad (36)$$

$$= \begin{bmatrix} A - 2j\delta & B \\ C - 2j\delta & D \end{bmatrix}$$

where the real $ABCD$ elements describe all the rest of the cavity except the gaussian aperture or VRM

mirror reflectivity. It is also convenient to define the parameters

$$N_{ga} \equiv \frac{w_{ga}^2}{B\lambda} \quad \text{and} \quad \delta \equiv \frac{B\lambda}{2\pi w_{ga}^2} = \frac{1}{2\pi N_{ga}} \quad (37)$$

where the quantity N_{ga} is a Fresnel number for the gaussian aperture, based on the $1/e$ width w_{ga} of the aperture and the effective roundtrip length B of the cavity. (This effective length B will in general not be the same as the real physical length of the cavity, because of the effects of lenses and other focusing elements within the cavity.) Note also that the central reflectivity value R_0 of the mirror is not included in these expressions, since the $ABCD$ analysis inherently only includes the quadratic variation about the axis, not the central value on the axis itself.

In resonators of practical interest, the gaussian aperture size w_{ga} will normally be at least somewhat larger than the minimum or confocal spot size $w_{cf} \equiv \sqrt{B\lambda/\pi}$ for the resonator in question. Hence, the Fresnel number N_{ga} will normally be greater than unity, typically by a considerable factor. This means we can in practical cases almost always make the approximation that $\delta \ll 1$. Note also that if the output mirror has a spherical curvature with radius of curvature R_{ga} , as well as a variable-reflectivity aperture size w_{ga} , its complex $ABCD$ matrix can be separated into the form

$$\begin{bmatrix} 1 & 0 \\ -2/R_{ga} - j\lambda/\pi w_{ga}^2 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ -2/R_{ga} & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ -j\lambda/\pi w_{ga}^2 & 1 \end{bmatrix} \quad (38)$$

and the real part of the matrix, representing the focusing power of the curved mirror, can be absorbed into the real $ABCD$ matrix for the rest of the round trip as shown in Equation (36). It will also be convenient to write the complex \tilde{m} value for the VRM resonator into the form

$$\tilde{m} \equiv \frac{\tilde{A} + \tilde{D}}{2} = m - j\delta \quad (39)$$

where the real part $m = (A + D)/2$ is the m -value for the real resonator without the variable-reflectivity mirror.

From Equation (23) the eigen- \tilde{q} -values for this

resonator are then given by

$$\begin{aligned} \frac{1}{\tilde{q}_a} \cdot \frac{1}{\tilde{q}_b} &= \frac{\tilde{D} - \tilde{A}}{2\tilde{B}} \mp \frac{\sqrt{\tilde{m}^2 - 1}}{\tilde{B}} \\ &= \frac{D - A}{2B} + j\frac{\delta}{B} \mp \frac{\sqrt{(m - j\delta)^2 - 1}}{B} \end{aligned} \quad (40)$$

As a practical matter we will most often be interested in resonators where the underlying resonator is geometrically unstable, i.e., $|m| > 1$. For simplicity, let us also limit the following discussion to positive-branch resonators, for which both the effective length B and the real half-trace parameter m are positive values. (The opposite case can easily be worked out, and the resulting answers are essentially the same as we will derive here.) We must then use the lower sign in Equation (40) for the confined and perturbation-stable mode. Within the approximation that $\delta \ll 1$, the confined and perturbation-stable eigenmode is found to have the gaussian \tilde{q} parameter

$$\begin{aligned} \frac{1}{\tilde{q}} &\equiv \frac{1}{R} - j\frac{\lambda}{\pi w^2} \approx \left[\frac{M - A}{B} \right] - j \left[\frac{2\delta}{(M^2 - 1)B} \right] \\ &\approx \left(\frac{M - A}{B} \right) - j\frac{1}{M^2 - 1} \times \left(\frac{\lambda}{\pi w_{ga}^2} \right) \end{aligned} \quad (41)$$

where $M \equiv m + \sqrt{m^2 - 1}$ is the round-trip geometric magnification of the underlying unstable resonator without the gaussian aperture, and where we have neglected terms of order δ^2 and higher in expanding the square root in Equation (40).

From the preceding equation, the spot size w_{int} of the lowest-order gaussian eigenmode for this resonator just inside the output mirror is given by

$$w_{int} \approx \sqrt{M^2 - 1} \times w_{ga} \quad (42)$$

and the radius of curvature for the output beam at the same plane (neglecting a small second-order correction) is given by

$$R_{int} \approx \frac{B}{M - A} \quad (43)$$

If the underlying resonator without the variable mirror reflectivity is a confocal unstable resonator, the real part of the resonator will have $A = M$, and hence $R_{int} = \infty$. The VRM output beam will then be automatically collimated (except, again, for the

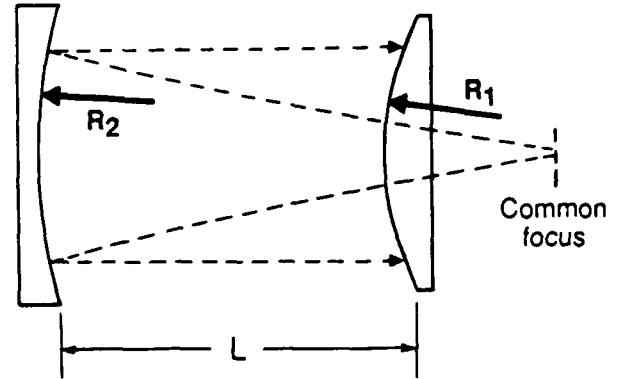


Fig. 5. Positive-branch confocal unstable resonator with magnification $M = 3$.

small second-order correction term). If not, the output mirror can be given the necessary lensing power to convert the curved wavefront of Equation (43) into a collimated beam after it passes through the output mirror.

After each round trip around the cavity, the complex amplitude of this lowest-order gaussian mode in the VRM resonator will be reduced by the complex eigenvalue

$$\tilde{\lambda}_b = \tilde{m} - \sqrt{\tilde{m}^2 - 1} \approx \frac{1}{M} + j\frac{2\delta}{M^2 - 1} \quad (44)$$

where we have again assumed a positive-branch unstable resonator with m and $B > 0$, and therefore again used the lower sign before the radical in Equation (33). Also, the approximation is again carried only to first order in the small quantity δ . We see that the matrix eigenvalue for the VRM resonator is essentially just the geometric eigenvalue $1/M$ with a very small correction due to the δ term.

4.3 Confocal Example

As one particularly simple and useful example of these formulas we can consider the confocal standing-wave VRM cavity shown in Figure 5. The mirror radii of curvature R_1 and R_2 and the cavity length L for this resonator must satisfy the confocal criteria

$$R_1 + R_2 = 2L \quad \text{and} \quad M = -R_2/R_1 \quad (45)$$

and hence

$$R_1 = \frac{-2L}{M - 1}, \quad R_2 = \frac{2ML}{M - 1} \quad (46)$$

The round-trip $\hat{A}\hat{B}\hat{C}\hat{D}$ matrix is then

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} = \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/R_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2/R_1 - j\lambda/\pi w_{ga}^2 & 1 \end{bmatrix} \quad (47a)$$

with the results that

$$\begin{aligned} \hat{A} &= M - j(M+1)L\lambda/M\pi w_{ga}^2 \\ \hat{B} &= (M+1)L/M \\ \hat{C} &= -jL\lambda/M\pi w_{ga}^2 \\ \hat{D} &= 1/M \end{aligned} \quad (47b)$$

Hence the relevant Fresnel number and \tilde{m} parameters in this case are

$$\begin{aligned} N_{ga} &= \frac{M}{M+1} \frac{w_{ga}^2}{L\lambda} \\ \text{and} \\ \tilde{m} &= \frac{M^2+1}{2M} - j \frac{M+1}{M} \frac{L\lambda}{2\pi w_{ga}^2} \end{aligned} \quad (49)$$

and the confined and self-consistent gaussian \tilde{q} parameter is

$$\frac{1}{\tilde{q}_{int}} \approx -j \frac{\lambda}{\pi(M^2-1)w_{ga}^2} \approx -j \frac{\lambda}{\pi w_{int}^2} \quad (50)$$

This says that the lowest-order gaussian mode inside the VRM resonator is a gaussian beam with a nearly planar radius of curvature

$$R_{int} \approx \infty \quad (51)$$

and a spot size w_{int} just inside the output plane given by Equation (42). Figure 4 illustrates the general shape of the lowest-order gaussian beam elsewhere inside this resonator.

4.4 VRM Resonator Design Criteria

Equations (34) through (44) contain all the basic analytical results needed for the design of a positive-branch geometrically unstable VRM resonator with Fresnel number N_{ga} of a few times unity or greater. They predict a gaussian beam inside the resonator with a gaussian spot size given by Equation (42), an output radius of curvature given by Equation (43), and a perturbation eigenvalue given by Equation (44).

One must be cautious in interpreting these results for real resonators, however. First of all, the perturbation or matrix eigenvalue given in Equation (44) takes into account the gaussian variation in one transverse direction only, and it does not include the finite reflectivity factor $R_0^{1/2}$ for E field or R_0 for optical intensity which multiplies the gaussian reflectivity variation in Equations (34) or (35). To include these factors, suppose we introduce the notations $\tilde{\gamma}_n$ to indicate the total reduction in mode amplitude, and $|\tilde{\gamma}_n|^2$ to indicate the total reduction in mode power, for an n -th order Hermite-gaussian eigenmode in a VRM resonator with a reflectivity variation in one transverse direction only (i.e., a strip resonator); and $\tilde{\gamma}_{nm}$ and $|\tilde{\gamma}_{nm}|^2$ to indicate the corresponding quantities for the nm -th eigenmode in a VRM resonator with reflectivity variation in both transverse directions. The fractional power loss per round trip in each case will then be given by $L_n \equiv 1 - |\tilde{\gamma}_n|^2$ or $L_{nm} \equiv 1 - |\tilde{\gamma}_{nm}|^2$, respectively. For the one-dimensional or strip resonator case we will then have

$$\begin{aligned} L_n &\equiv 1 - |\tilde{\gamma}_n|^2 \\ &= 1 - R_0 |\tilde{\lambda}_n|^{2n+1} \approx 1 - R_0/M^{2n+1} \\ &\approx 1 - R_0/M \quad (\text{lowest-order mode}) \end{aligned} \quad (51)$$

while for the two-transverse-dimension case we will have

$$\begin{aligned} L_{nm} &\equiv 1 - |\tilde{\gamma}_{nm}|^2 \\ &= 1 - R_0 |\tilde{\lambda}_n|^{2n+1} |\tilde{\lambda}_m|^{2m+1} \\ &\approx 1 - R_0/M^{2(n+m+1)} \\ &\approx 1 - R_0/M^2 \quad (\text{lowest-order mode}) \end{aligned} \quad (52)$$

To put this in another way, the ratio of round-trip power reduction factors for the $n = 0$ and $n = 1$ modes in the strip resonator case will be

$$\frac{|\tilde{\gamma}_1|^2}{|\tilde{\gamma}_0|^2} = \frac{1}{M^2} \quad (\text{strip resonator}) \quad (53)$$

while for a circularly symmetric VRM resonator, the loss discrimination between lowest and next-lowest-order eigenmodes will be

$$\frac{|\tilde{\gamma}_{10}|^2}{|\tilde{\gamma}_{00}|^2} = \frac{|\tilde{\gamma}_{01}|^2}{|\tilde{\gamma}_{00}|^2} = \frac{1}{M^3} \quad (\text{circular resonator}) \quad (54)$$

These ratios determine the mode discrimination against higher-order modes in the VRM resonator. In general, for any value of geometric magnification M that is a small amount above unity these ratios provide very good transverse mode discrimination, at least compared to conventional stable optical resonators.

The second significant point in VRM resonator design is that the transverse profile of the lowest-order eigenmode is gaussian only *inside* the laser cavity. The output beam from the resonator will have this gaussian profile, multiplied by the radially varying mirror transmission $T(r) = 1 - R(r) = 1 - |\tilde{\rho}(r)|^2$ from Equation (34) or (35). (This assumes a lossless variable reflection technique, so that $R(r) + T(r) = 1$ everywhere.) Multiplying the radially decreasing gaussian profile of Equation (42) by the mirror transmission obtained from Equation (35) gives for the output beam profile in a circularly symmetric resonator

$$\begin{aligned} I(r) &= [1 - R(r)] \times \exp[-2r^2/w_{int}^2] \\ &= \exp\left[-\frac{2r^2}{w_{int}^2}\right] - R_0 \exp\left[-\frac{2M^2 r^2}{w_{int}^2}\right] \end{aligned} \quad (55)$$

where $w_{int}^2 \approx (M^2 - 1)w_{ga}^2$ is the value inside the resonator.

Figure 6 shows typical plots of the output mirror transmission profile, the gaussian beam profile just inside the output mirror, and the output beam profile (normalized to the same peak value) just outside the mirror, for the particular case of $M = \sqrt{2}$ and several different choices of central mirror reflectivity R_0 . To have an output beam profile without a null or a significant dip in the center, the output mirror reflectivity R_0 at the center must be less than unity, so that there is some finite transmission through the mirror even at the center. The most uniform output beam profile will be obtained by setting the central reflectivity at the maximally flat condition where the central dip is just about to appear. One can show from Equation (55) that this maximally flat condition is given by

$$R_0[\text{maximally flat}] \equiv R_{MF} = \frac{1}{M^2} \quad (56)$$

Hence the effective output coupling for the lowest-order gaussian mode in the maximally flat case is given by

$$L_{00}[\text{maximally flat}] = 1 - \frac{R_{MF}}{M^2} = 1 - \frac{1}{M^4} \quad (57)$$

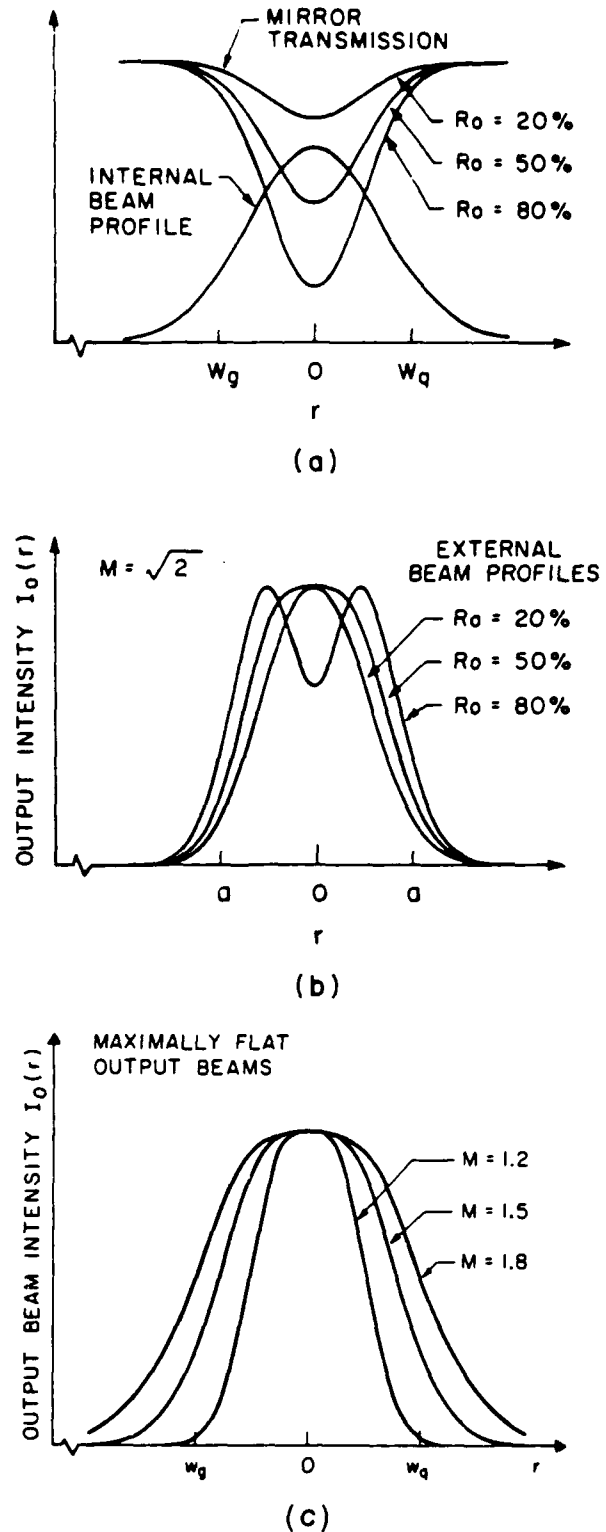


Fig. 6. Internal beam profile and output mirror transmission profiles for different central mirror reflectivities. (b) Resulting output beam profiles, outside the variable-reflectivity mirror. (c) Maximally flat output beam profiles for different geometric magnifications M .

where this formula applies for either strip or circular resonators.

The elementary design tradeoffs for a geometrically unstable VRM resonator can thus be outlined as follows:

1) Select a value of geometric magnification M which will give an adequate mode intensity discrimination ratio between the lowest and next higher-order transverse eigenmodes, as described by Equations (51) to (54).

2) Select a value of mirror-center reflectivity R_0 which satisfies if possible the maximally flat criterion $R_{MF} = 1/M^2$, hoping that the resulting output coupling $L_{00} = 1 - 1/M^4$ for the lowest-order mode will not overcouple the oscillator.

3) Finally, choose the width parameter w_{ga} of the gaussian mirror so that the gaussian spot size $w_{int} \approx \sqrt{M^2 - 1} \times w_{ga}$ adequately fills the gain medium inside the resonator.

In a low-gain laser system the maximally flat value of R_0 obtained in step 2 may overload the oscillator. If so, one will have to drop back to a smaller value of magnification in step 1, or accept a larger value of R_0 in step 2 (which will lead to a central dip in the beam profile), until the lowest-order mode coupling given by Equations (51) or (52) becomes acceptable for the laser medium in question. If the mirror-center reflectivity is larger than the maximally flat value R_{MF} given by Equation (56), one can show that the intensity reduction at the center of the beam ($r = 0$) compared to the value at the peak of the resulting "donut" profile (at $r = r_p$) is given by

$$\frac{I(0)}{I(r_p)} = \frac{(1 - R_0)R_0^{1/(M^2-1)}(M^2)^{M^2/(M^2-1)}}{M^2 - 1} \quad (58)$$

where $r_p = [\ln(MR_0^{1/2})]^{1/2} \times w_{ga}$.

4.5 Other VRM Resonator Properties

There are a number of additional properties and analytical aspects of VRM resonators which we can only briefly mention here, but which are discussed in more detail in the References listed at the end of this tutorial. These include:

1. Higher-order eigenmodes

The lowest-order or $n = 0$ and $n = 1$ modes in the gaussian VRM resonator are both physically realizable and perturbation stable, since they represent

confined modes for which the perturbation eigenvalue has magnitude $|\lambda_b| < 1$. But, this automatically means that all higher-order Hermite-gaussian modes (with indices $n \geq 2$) will be perturbation unstable, since they will have a growth rate proportion to $|1/\lambda_b|$ for perturbations $\Delta\bar{v}$ to the self-consistent Hermite scale factor \bar{v} . Therefore, oscillation in higher-order Hermite-gaussian modes should not be sustainable in an ideal gaussian VRM laser cavity.

This can be understood in physical terms as follows. Suppose one starts with some higher-order Hermite-gaussian mode of order $n = N \geq 2$ circulating in such a resonator and then perturbs this field pattern in some fashion. The resulting field pattern can then be reexpanded, at least conceptually, as a sum over the unperturbed Hermite-gaussian modes of all orders, i.e., including $n = N \pm 1$, $N \pm 2$, and so on. In fact, if one perturbs the \bar{v} value for an N -th order eigenmode, this expansion will include components (possibly quite small) extending all the way down to the $n = 1$ or $n = 0$ modes. As the perturbed field then propagates repeatedly around the cavity, all the higher-order modes with $n \geq 1$ will, because of their higher losses, begin to die out relative to the $n = 0$ mode. After a large enough number of round trips this will eventually leave only the $n = 0$ mode remaining as the dominant mode in the cavity.

The important point is thus not so much that the higher-order modes are "perturbation unstable" as that they have high losses and will necessarily die out compared to the $n = 0$ mode. Of course if one deliberately introduces some lossy perturbing element inside the VRM cavity—for example, some distribution of tiny point scatterers or wires or lossy dots—it is entirely possible that one can change the relative mode losses in the cavity so as to force the perturbed cavity to oscillate at steady-state in a mode identical to, or very close to, one of the higher-order Hermite-gaussian modes, whichever one has a pattern which minimizes the impact of these lossy scatterers.

2. Far-field beam patterns

The output beam from a VRM resonator, just outside the variable-reflectivity mirror, is not a simple gaussian profile, and hence neither is the far-field beam profile. Because of the generally smooth and uniphase character of the near-field profile, however, the far-field beam will in general have very good beam quality. The far-field beam profile for the out-

put beam given by Equation (55) can be expanded in a series which has the form of a dominant gaussian beam plus a series of gaussian holes of decreasing amplitude. Exact calculation of the far-field beam pattern for the VRM resonator is thus not a particularly difficult problem, as described in some of the References.

3. Misalignment sensitivity

The first-order effect of a misalignment in a VRM resonator is to couple lowest and higher-order gaussian modes so that the eigenmodes become admixtures of the usual Hermite-gaussian modes. The resulting analysis is somewhat complicated but straightforward. In general, the conclusion is that VRM resonators should be insensitive to small angular misalignments of the mirrors.

4. Radially variable gain media

A radially varying gain medium, in which the laser gain has a maximum on axis and decreases off axis, can also play the same role as a variable-reflectivity mirror or variable-transmission aperture, with the gain medium being treated as a "gaussian duct" and described by a complex-valued $\tilde{A}\tilde{B}\tilde{C}\tilde{D}$ ray matrix. The resulting eigenmodes in the resonator will again be confined and perturbation-stable Hermite-gaussian modes, given by essentially the same analysis as in this tutorial. Certain lasers do operate more or less in this fashion.

By the same token, a medium in which the laser gain increases moving radially outward from the optical axis should, at least in theory, produce a set of eigenmodes which are either perturbation-stable but unconfined, or else perturbation-unstable but confined, as in Subsection 4 of Section 3.5 above. Some laser media do in fact exhibit radially increasing gain profiles at least near the central axis, due for example to pump light absorption, or to excess heating on axis in gas lasers. In these cases, however, first of all the gain increase usually stops at some finite radius, and second the optical mode itself is usually truncated or confined by finite apertures within the laser cavity. Despite their theoretical instability, therefore, laser media with radially increasing gain can usually be used in practical lasers—although not with the advantageous properties described for VRM resonator lasers in this tutorial.

5. Alternative mirror reflectivity profiles

The gaussian mirror reflectivity profile of Equa-

tions (34) and (35), though by far the simplest to analyze, may not be quite the optimum profile for a VRM resonator, and also this expression may not match accurately to the real reflectivity profile of a real variable-reflectivity mirror. Other reflectivity profiles have also been analyzed, to varying degrees of approximation, as discussed in the References. In particular, the "super-gaussian" profile

$$R(r) = R_0 \exp[-2(r/w_{ga})^n] \quad (59)$$

where n is some integer greater than 2, and various forms of quadratic and higher-order profiles of the form

$$R(r) = R_0 [1 - 2(r/w_{ga})^n] \quad (60)$$

have been considered.

In practical situations one has also to consider the effect of limiting hard-edged apertures which are likely to be present due to finite tube diameters, finite mirror edges, or other apertures inside the cavity. Some of these effects are considered in the References. From experience with diffraction effects in other cavities one can estimate that truncating a circular optical beam with a hard circular aperture at the point where the beam intensity is down by a factor of δ from the peak intensity value in the near field, will produce near-field intensity ripples with peak variation of $\pm 2\delta$ in intensity, and a peak on-axis intensity reduction possibly as large as -2δ in the far field. That is, truncating a smooth circular beam at the 1% intensity radius in the near field can produce $\pm 20\%$ near-field Fresnel intensity ripples, and a 20% reduction in the far-field intensity on axis.

V. PRACTICAL VARIABLE-REFLECTIVITY MIRRORS

Finally, given the attractive properties of VRM resonators for laser applications, we can observe that practical variable-reflectivity mirrors or coupling apertures which can be used in VRM laser resonators will obviously be potentially very useful optical components. While the technology of such elements still needs further development, examples of variable-reflectivity mirrors or couplers which have been demonstrated to date include:

1. Tapered-Reflectivity Dielectric Mirrors.

Lossless output mirrors with radially varying reflectivity can be made by evaporating multilayer dielectric coatings which are suitably tapered in the radial

direction onto a transparent substrate. Such tapered dielectric mirrors, if they became readily available, could be very useful and effective; and initial experiments with such tapered mirror coatings have been notably successful (see References). Such coatings may still be somewhat difficult and expensive to make, and of course the radial reflectivity variation is fixed and cannot be changed in operation. One must also be concerned with possible radially varying phase distortion of the reflected or transmitted beams due to the radially varying coatings. Where mirrors of this type are available, however, they are probably the method of choice for a practical VRM laser.

2. *Radially Variable Absorbing Apertures.* Apertures with radially varying absorption or transmission properties can also be fabricated using evaporated metal coatings, exposed and developed photographic films, or absorptive filters with radially tapered density. Aside from the difficulties of fabricating such apertures, they have the practical disadvantage that they absorb power and hence reduce the laser power efficiency, and they may also be damage-prone in high power lasers. Their value, if any, is probably only for test experiments, not practical laser devices.

3. *Radially Tapered Diffraction Gratings.* A somewhat expensive but effective type of variable-reflectivity mirror or output coupler can be obtained by fabricating a diffraction grating in which the grating period (and hence the diffraction angle) remains constant across the grating, but in which the groove depth (and hence the diffraction efficiency) is tapered from the center to the edge of the grating.

This type of grating can then be used with one diffraction order (perhaps the specular reflection, or the back-reflected beam) providing the feedback beam in the laser resonator, and another order providing the radially varying coupling. Depending on how the grating is used, the grating strength may be required to increase or decrease with distance from the grating center. Applications to either standing-wave or ring cavities are easily envisioned.

A variable-depth grating of this type is obviously somewhat difficult to fabricate, but once fabricated could be efficient and effective. At least one metal grating mirror of this type has been prepared by ion beam milling techniques and used with success in a high-power CO₂ laser. Such a metal grating mirror can, of course, be internally water-cooled at

high power densities.

4. *Radially Varying Birefringent Couplers.* Another innovative and effective solution to the variable-reflectivity mirror problem is the radially variable birefringent coupling technique developed by Byer and co-workers (see References). In this technique a birefringent element (or, in a ring cavity, an optically active element) with a radially varying thickness is placed inside the laser cavity. The wave passing through the element will have its polarization rotated, or converted from linear to elliptical, by an amount which increases with radius from the center of the element. A polarization-sensitive coupling element, such as a polarizing crystal, a Brewster-angle plate, or a dielectric beam splitter with dichroic properties, is then used to couple out the rotated polarization component with a coupling strength that increases radially at the desired rate. In more advanced versions two polarization-rotating elements of opposite sign, shaped like positive and negative lenses, can be used in cascade to double the radially varying birefringence while cancelling the ordinary focusing effects produced by each element alone.

One of the advantages of this design is that by changing or rotating the different elements involved (the polarizing elements and the output coupling elements) one can obtain many different radial profiles, and can even vary the radial coupling profile in operation. In wavelength regions where good-quality low-loss polarizing elements can be obtained, this approach, though slightly complex, can be effective, flexible, and can withstand at least moderate peak and average optical power levels.

5. *Apodized Interferometers, or "Newton's Ring" Couplers.* A radially varying amplitude reflection and transmission pattern can also be produced by the combination of a flat and a curved surface placed in contact, or in close proximity, as will be evident to anyone who has ever observed a Newton's ring interference pattern. The central fringe of such a Newton's ring pattern, depending on whether it is light or dark in reflection or transmission, can provide either a radially varying mirror for a standing-wave cavity or a radially varying coupling element for a ring cavity. Such an element is potentially simple and inexpensive; and by controlling the curvatures, individual surface reflectivities, and spacing of the two elements (perhaps with a piezoelectric transducer), it is possible to control the

central reflectivity and the radial variation of the element independently. This approach thus looks very promising and has been explored with success by several groups (see References).

VI. CONCLUSIONS

The variable-reflectivity resonator appears overall to be a very promising technique for combining the advantages of the gaussian stable and geometrically unstable resonator approaches, and as such should be more widely developed in the future. Finding low-cost, low-loss, and low-damage methods for generating the necessary variable reflectivity mirrors appears to be the primary problem at the minute. Several of the techniques described in the preceding section appear to have considerable future potential at least for low to medium-power lasers in the visible and near infrared. The necessary elements can be relatively simple to fabricate and adjust, can have low insertion losses, and should have relatively high optical damage thresholds, at least in the visible region where suitable optical materials are available. Finding suitable optical elements for high average power lasers will be somewhat more difficult, especially in the infrared. New alternative solutions to the general variable-reflectivity coupler problem with similarly useful properties are thus well worth inventing.

VII. ACKNOWLEDGMENTS

This work was sponsored by the Air Force Office of Scientific Research. The author appreciates the kindness of a number of authors who sent him preprints of works cited in the References, and also the hospitality of the Max-Planck Institute for Quantum Electronics in Garching (Munich), Germany, where parts of this manuscript were prepared.

REFERENCES

General References

- A. E. Siegman, *LASERS* (University Science Books, Mill Valley, California, 1986); Section 23.3.
J. A. Arnaud, "Optical resonators in the approximation of Gauss," *Proc. IEEE* **62**, 1561-1570 (November 1974).

Unstable Resonators with Tapered Mirror Reflectivity

- A. Anan'ev and V. E. Sherstobitov, "Influence of the edge effects on the properties of unstable resonators," *Soviet J. Quantum Electron.* **1**, 263-267 (November-December 1971).
V. E. Sherstobitov and G. N. Vinokurov, "Properties of unstable resonators with large equivalent Fresnel numbers," *Soviet J. Quantum Electron.* **2**, 224-229 (Nov.-Dec. 1972).
G. L. McAllister, W. H. Steier, and W. B. Lacina, "Improved mode properties of unstable resonators with tapered reflectivity mirrors and shaped apertures," *IEEE J. Quantum Electron.* **QE-10**, 346-355 (March 1974).
E. A. Maunders, G. L. McAllister, and W. H. Steier, "Experiments on improved unstable resonator mode profiles by aperture shaping," *IEEE J. Quantum Electron.* **QE-10**, 821-822 (October 1974).
N. Kumagai, "Resonant modes in a Fabry-Perot resonator consisting of nonuniform reflectors," *Elec. and Comm. Japan* **49**, 1-8 (July 1966).

Early Gaussian VRM Analyses

- S. N. Vlasov and V. I. Talanov, "Selection of axial modes in open resonators," *Radio Engineering and Electronic Physics* **10**, 469-470 (March 1965).
N. G. Vakhimov, "Open resonators with mirrors having variable reflection coefficients," *Radio Engineering and Electronic Physics* **10**, 1439-1446 (September 1965).
H. Zucker, "Optical resonator with variable reflectivity mirrors," *Bell System Technical Journal* **49**, 2349-2376 (November 1970).
A. N. Chester, "Mode selectivity and mirror misalignment effects in unstable laser resonators," *Applied Optics* **11**, 2584-2590 (November 1972).
Y. Suematsu, et al., "A light beam waveguide using Gaussian mode filters," *Elec. and Comm. in Japan* **51-B**, 67-74 (April 1968).

Resonator Analysis Using Complex ABCD Matrices

- L. W. Casperson, "Mode stability of lasers and periodic optical systems," *IEEE J. Quantum Electron.* **QE-10**, 629-634 (September 1974).

Gaussian VRM Analysis Using Complex ABCD Matrices

L. W. Casperson and Susan D. Lunnam, "Gaussian modes in high loss laser resonators," *Applied Optics* 14, 1193-1199 (May 1975).

A. Yariv and P. Yeh, "Confinement and stability in optical resonators employing mirrors with gaussian reflectivity tapers," *Optics Communications* 13, 370-374 (April 1975).

U. Ganiel and A. Hardy, "Eigenmodes of optical resonators with mirrors having Gaussian reflectivity profiles," *Applied Optics* 15, 2145-2149 (September 1976).

Gain-Guided Resonators

U. Ganiel and Y. Silberberg, "Stability of optical resonators with an active medium," *Applied Optics* 14, 306-309 (February 1975).

U. Ganiel, A. Hardy, and Y. Silberberg, "Stability of optical laser resonators with mirrors of gaussian reflectivity profiles, which contain an active medium," *Optics Communications* 14, 290-293 (July 1975).

A. Hardy, "Gaussian modes of resonators which contain saturable gain medium," private communication.

Experiments on VRM Resonators

P. Lavigne, A. Parent, D. Pascale, and N. McCarthy, "A compact wide-aperture single-mode CO₂ laser with a low chirp rate," *IEEE J. Quantum Electron.* QE-22, 2200-2203 (December 1986).

N. McCarthy and P. Lavigne, "Large-size Gaussian mode in unstable resonators using Gaussian mirrors," *Optics Letters* 10, 553-555 (November 1985). (Excellent results!)

Misalignment Sensitivity

N. McCarthy and P. Lavigne, "Optical resonators with Gaussian reflectivity mirrors: misalignment sensitivity," *Applied Optics* 22, 2704-2708 (1 September 1983).

Far-Field Beam Characteristics

N. McCarthy and P. Lavigne, "Optical resonators with Gaussian reflectivity mirrors: output beam characteristics," *Applied Optics* 23, 3845-3850 (1 November 1984).

VRM Resonators: Nongaussian Apertures

A. Parent, N. McCarthy, and P. Lavigne, "Effects of limiting hard apertures in optical resonators with Gaussian reflectivity mirrors," 1986 International Conference on Quantum Electronics (May 1986).

A. Parent, N. McCarthy, and P. Lavigne, "Effects of hard apertures on mode properties of resonators with Gaussian reflectivity mirrors," *IEEE J. Quantum Electron.* QE-23, 222-228 (February 1987).

P. Lavigne, A. Parent, D. Pascale and N. McCarthy, "A wide-aperture single-mode compact TE-CO₂ laser with a low chirp rate," 1986 CLEO (May 1986).

A. Parent, P. Lavigne, "Laser resonators with graded-reflectivity mirrors," (abstract only), Paper W139, 1986 CLEO Meeting.

S. De Silvestri, P. Laporta, V. Magni, and O. Svelto, "Unstable laser resonators with super-Gaussian mirrors," *Opt. Lett.* 13, 201-203 (March 1988).

Radially Tapered Dielectric Mirrors

P. Lavigne, N. McCarthy, and J-G. Demers, "Design and characterization of complementary Gaussian reflectivity mirrors," *Applied Optics* 24, 2581-2586 (15 August 1985).

C. Zizzo, C. Arnone, C. Cali, and S. Sciortino, "Fabrication and characterization of tuned Gaussian mirrors for the visible and the near infrared," *Optics Letters* 13, 342-344 (May 1988).

Apoditic Interferometers and Newton's Ring Mirrors

E. Armandillo and G. Giuliani, "Achievement of large-sized TEM₀₀ mode from an excimer laser by means of a novel apoditic filter," *Optics Letters* 10, 445-447 (September 1985).

S. De Silvestri, P. Laporta, V. Magni, and O. Svelto, "Radially variable reflectivity output coupler of novel design for unstable resonators," *Optics Letters* 12, 84-86 (February 1987).

S. De Silvestri, P. Laporta, and V. Magni, "Laser output coupler based on a radially variable interferometer," *J. Opt. Soc. Am. A* 4, 1413-1418 (August 1987).

Radially Varying Birefringent Couplers

J. M. Eggleston, G. Giuliani, and R. L. Byer, "Radial intensity filters using radial birefringent elements," *J. Opt. Soc. Am.* 71, 1264-1272 (October 1981).

G. Giuliani, Y. K. Park, and R. L. Byer, "Radial birefringent element and its application to laser resonator design," *Optics Letters* 5, 491-493 (November 1980).

D. J. Harter and J. G. Walling, "Low-magnification unstable resonators used with ruby and alexandrite lasers," *Optics Letters* 11, 706-708 (November 1986).

The Virtual-Source Theory for Unstable Resonators:
A Reformulation and Some New Results

Paul Mussche, and A. E. Siegman
Edward L. Ginzton Laboratory
Stanford University

I. INTRODUCTION

Unstable optical resonators have ever since their initial discovery [Siegman, 1965] resisted simple analytical treatment. The purely geometrical approach to unstable resonator theory does provide a remarkably good "zero-order" explanation for the basic geometrical properties of these resonators. Almost all the more advanced properties of unstable resonator modes have been determined, however, by means of laborious numerical calculations [xxx]. Detailed analytical treatments of the unstable resonator have been carried out using several different physical models [xxx]. The resulting analyses have generally been very complex, however, and obtaining quantitative results from these analyses has generally required as much or more numerical work than a straightforward Fox-and-Li type purely numerical analysis. Attempts at finding simple extensions of the purely geometrical theory to take diffraction effects into account to a first-order or perturbative approximation have generally run into fundamental mathematical difficulties, and no good set of basis modes for expanding unstable resonators has been found.

The one exception to this picture is an approach pioneered by Horwitz and referred to by him as the *asymptotic approach* [Horwitz, 1973, 1976; Moore and McCarthy, 1977; Butts and Avizonis, 1978], or by Southwell as the *virtual source approach* [Southwell, 1981, 1986]. This approach does permit the detailed mode properties of either strip or circular-mirror unstable resonators to be calculated with only a small amount of numerical effort, both in the low Fresnel number region and especially in the large Fresnel number region where straightforward Fox-and-Li calculations become prohibitively expensive.

The formulation of the Horwitz asymptotic theory remains somewhat complicated, however, and its mathematical appearance rather forbidding, despite its obvious effectiveness. One objective of this paper is to follow Southwell's lead [Southwell, 1981] and present a reformulation of this theory which emphasizes the physical basis of the theory in the form of diffracted edge waves from multiple apertures in an unstable lensguide. For reasons to be explained later, we will carry out this reformulation using a simple symmetric single-lens model for the unstable lensguide, rather than the "canonical model" employed by Horwitz and later workers. While there is really no new physics in our approach over and above the earlier derivations, our reformulation makes the physical basis of the analysis clearer, and allows us to obtain the primary results of the analysis in just a few lines.

As an application of this, we then use the virtual source theory to calculate a number of results verifying the theoretically predicted biorthogonality properties of unstable resonator modes. We also use these results to calculate numerical values of the "excessquantum noise factor" which is predicted to apply for unstable optical resonators.

The outline of this paper is as follows. In Section II we develop the physical basis of the virtual source theory and show how it applies to two alternative models for an unstable resonator, and in Section III we derive the basic polynomial equation that results from the virtual-source theory. In Section IV we present a number of eigenvalue results for typical resonator cases, and in Section V we confirm numerically that the resonator eigenmodes are biorthogonal to a set of transposed or reverse-traveling resonator modes, as predicted theoretically. We also consider the mode mixture that results when one attempts to excite the lowest-order eigenmode with maximum amplitude at the input to an unstable resonator or lensguide. Finally, in Section VI we discuss the excess noise factor that is predicted for spontaneous emission into unstable resonator modes because of their lack of conventional power orthogonality properties. Detailed formulas for the edge-wave functions in the strip-resonator and circular-mirror case are summarized in an Appendix.

II. DERIVATION OF THE VIRTUAL SOURCE THEORY

A. Basic Concepts of the Virtual Source Analysis

The repeated round trips of an eigenwave in an unstable (or stable) optical resonator can equally well be viewed as propagation of the same eigenwave through repeated sections of an equivalent iterated periodic lensguide; and adopting this viewpoint can in many cases simplify the analysis of resonator mode properties. A simple unstable resonator with a single output mirror can in particular be replaced either by a simple symmetric lensguide consisting of diverging lenses of focal length f spaced by distance L , as shown in Fig. 1(a), or by the somewhat more abstract canonical model shown in Fig. 1(b), consisting of a telescope of transverse magnification M and zero length and a free space section of length MB , where B is the B element of the $ABCD$ matrix from aperture to aperture in the actual unstable resonator. These two models are entirely equivalent, and can represent any unstable resonator having arbitrary intracavity paraxial elements plus a single output coupling mirror, since the properties of a single-aperture unstable resonator or lensguide depend only on the magnification M and equivalent Fresnel number F_{eq} of the system, and not on any of the detailed aspects of any of the internal lenses or mirrors encountered in a round trip around the cavity.

To introduce the virtual source concept, suppose we position our eye at some reference plane inside an unstable lensguide such as in Fig. 1, and look backward down the lensguide through the nearest aperture. In an unstable lensguide, if one looks back from any position on the optical axis of the system, one can see all of the earlier apertures of the lensguide. Each earlier aperture is, however, seen through an increasing number of lenses and any other optical elements contained within each period of the lensguide. As a result, each earlier aperture will appear to have a magnified (or demagnified) apparent size, and to be located at a different apparent distance back down the line. Considering each of these earlier apertures to be a potential radiation source for the output beam propagating along the unstable system is the essence of the virtual source approach.

Within the usual Fresnel degree of approximation, in fact, an optical wave passing through any finite aperture will produce a near-field pattern beyond that aperture which can be described to a good first approximation as a geometrical wave transmitted through the aperture, plus an edge-wave component which appears to radiate, or to be scattered, from the edges of the aperture.

FIGURE 1. (a) Symmetric and (b) canonical forms for a general one-aperture unstable resonator or lensguide.

The aperture edges in an unstable lensguide in particular are all strongly illuminated by the diverging eigenwave propagating along the lensguide. The basic concept of the virtual source theory, therefore, as we will show in the following sections, is that each eigenmode function, or eigenwave, at any reference plane in an unstable lensguide can be viewed as the summation of a series of edge-wave functions coming from the edges of all of the apertures seen looking back down the lensguide.

In practice, this means that the total field at any transverse reference plane within the lensguide can be viewed as a series of distinct edge waves coming from the edges of the dozen or so nearest apertures located the closest behind the reference plane, plus a spherical or geometric component which comes from a superposition of all the more distant apertures. The set of edge wave functions coming from the nearer apertures provides, so to speak, the most effective set of basis functions to use as perturbations on the geometric solution coming from the more distant apertures for expanding the complicated eigenwaves of an unstable system. Expanding the eigenmode field at any reference plane in terms of these edge waves, plus the geometric term, then leads to both the most compact mathematical expansion and to a simple polynomial eigenequation for the resonator or lensguide eigenvalues, as we will derive below.

B. Virtual Aperture Locations

Our first task is thus to develop formulas for the apparent sizes and distances of the earlier apertures in an unstable resonator system as seen looking back down the lensguide, as well as formulas for the edge wave functions diffracted from these earlier apertures.

The apparent positions of the aperture edges can be calculated from a purely geometrical ray viewpoint, using an $ABCD$ matrix formalism. Let us consider, for example, all those rays emerging with different slopes x'_1 from a source point having a transverse displacement x_1 at an

FIGURE 2. A source point x_1 as seen through an arbitrary $ABCD$ system appears to be transformed to transverse displacement x_1/D at distance B/D .

input plane to an $ABCD$ system, as illustrated in Fig. 2(a). The ray displacements x_2 and slopes x'_2 at the output plane from the $ABCD$ system will then be given by

$$\begin{aligned} x_2 &= Ax_1 + Bx'_1 \\ x'_2 &= Cx_1 + Dx'_1 \end{aligned} \quad (1)$$

Eliminating the input slope between these equations gives

$$x_2 - (B/D)x'_2 = D^{-1}x_1 \quad (2)$$

But, this says that these rays appear to emanate from an apparent source point, as seen through the $ABCD$ system, that is displaced off the axis by a displacement x_1/D and is located at a distance B/D behind the input plane.

To extend this argument to wave optics, we can note that the Huygens-Fresnel integral in one transverse dimension for propagation through a distance L in free space is

$$u_2(x_2) = \sqrt{\frac{j}{L\lambda}} \int u_1(x_1) \exp \left[-j \frac{\pi}{L\lambda} (x_1 - x_2)^2 \right] dx_1 \quad (3)$$

whereas the same integral for propagation through an arbitrary $ABCD$ system is

$$u_2(x_2) = \sqrt{\frac{j}{B\lambda}} \int u_1(x_1) \exp \left[-j \frac{\pi}{B\lambda} (Ax_1^2 - 2x_1x_2 + Dx_2^2) \right] dx_1 \quad (4)$$

If we make the changes of variables $y_1 = x_1/D$ and $L' \equiv B/D$, Eq. (4) can be cast into the

FIGURE 3. Looking backward to the N -th aperture in a symmetric unstable lensguide.

free-space form

$$u_2(x_2) = \sqrt{\frac{j}{L'\lambda}} \int v_1(y_1) \exp\left[-j \frac{\pi}{L'\lambda} (x_1 - y_1)^2\right] dy_1 \quad (5)$$

where the effective input wave is given by

$$v_1(y_1) \equiv D^{1/2} \exp\left[-j \frac{\pi CD}{\lambda} y_1^2\right] u_1(Dy_1) \quad (6)$$

This says again that each source point x_1 at the input to the $ABCD$ system looks as if it has been moved to an effective source point located at $y_1 = x_1/D$ off the axis and at a free-space distance $L \equiv B/D$ behind the input plane. This source field is also scaled by a factor $D^{1/2}$ (for one transverse dimension), and rotated in phase by the amount $\exp[-j\pi Cx_1^2/D\lambda]$.

C. Virtual Apertures For a Symmetric Unstable Lensguide

To apply the virtual-source approach to the symmetric lensguide of Fig. 1(a), let us first pick the reference plane for our analysis to be the end plane just before the output aperture of one section of the symmetric lensguide of Fig. 1(a). This means we will be calculating the fields at a plane just before striking the output mirror or output coupler in a real unstable resonator. Standing at this plane, we look backward from this reference plane down the lensguide toward the N -th aperture back, as shown in Fig. 3, and attempt to evaluate the overall $ABCD$ matrix coming forward from this N -th aperture to our reference plane.

We can first note that the $ABCD$ matrix for one complete period of the symmetric system, going from a midplane halfway between the lenses to a midplane one period later, because of

symmetry can be written in the form

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} m & (m+1)L/2 \\ 2(m-1)/L & m \end{bmatrix} \equiv \begin{bmatrix} \cosh \theta & Z_0 \sinh \theta \\ Z_0^{-1} \sinh \theta & \cosh \theta \end{bmatrix} \quad (7)$$

where $m \equiv (A + D)/2 \equiv \cosh \theta$ is the half-trace of this $ABCD$ matrix. For the example of Fig. 1(a), m has the value $m = 1 + L/2|f|$. The geometric magnification M per period of the positive-branch unstable system is then given by

$$M = m + \sqrt{m^2 - 1} = \cosh \theta + \sinh \theta = \exp(+\theta) \quad (8)$$

and the "characteristic impedance" Z_0 by

$$Z_0 = \sqrt{\frac{m+1}{m-1}} \times \frac{L}{2} = \frac{M+1}{M-1} \times \frac{L}{2} = \frac{L}{2 \tanh(\theta/2)} \quad (9)$$

This characteristic impedance turns out to be just the radius of curvature of the geometrical wave at the midplane of the unstable resonator. The transformation of the $ABCD$ matrix of Eq. (7) through N cascaded midplane-to-midplane sections in sequence can then be easily written as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = \begin{bmatrix} \cosh N\theta & Z_0 \sinh N\theta \\ Z_0^{-1} \sinh N\theta & \cosh N\theta \end{bmatrix} \quad (10)$$

i.e., the matrix of Eq. (7) raised to the N -th power.

The total ray or wave propagation from the N -th aperture back, propagating forward to the reference plane, then consists in sequence of the lens immediately in front of the N -th aperture back; free-space propagation through length $L/2$ from this lens to the next nearest midplane; $N - 1$ midplane-to-midplane sections evaluated according to the formula in Eq. (10); and finally an additional half-period of free-space propagation from the final midplane to the reference plane. Multiplying the $ABCD$ matrices for these sequential elements (in reversed order) then gives for the total $ABCD$ matrix from the N -th aperture back, all the way up to the reference plane, of

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cosh(N + 1/2)\theta / \cosh \theta/2 & L \sinh N\theta / \sinh \theta \\ 2 \tanh(\theta/2) \sinh N\theta / L & \cosh(N - 1/2)\theta / \cosh \theta/2 \end{bmatrix} \quad (11)$$

The matrix elements can also be manipulated into various other hyperbolic forms using standard hyperbolic identities.

Using these $ABCD$ values to determine the effective size and distance of the N -th aperture back, as discussed in Eqs. (2) through (6), shows that each aperture has a successively smaller halfwidth a_N given by

$$a_N = \frac{a}{D} = \frac{a \cosh \theta/2}{\cosh(N - 1/2)\theta} \quad (12)$$

and each aperture appears to lie at a distance behind the reference plane given by

$$L_N = \frac{B}{D} = \frac{L}{2} \left[\frac{\tanh(N - 1/2)\theta}{\tanh \theta/2} + 1 \right]. \quad (13)$$

FIGURE 4. Apparent apertures seen looking backward down a symmetric unstable lensguide.

Figure 4 illustrates the apparent sizes and locations of the earlier apertures for a typical symmetricunstable resonator case.

Note in particular that the apparent distances for all the more distant apertures converge to a limiting value given by

$$L_{\infty} \equiv \lim_{N \rightarrow \infty} \frac{B}{D} = \frac{M}{M-1} L = R_g \quad (14)$$

where R_g is the radius of curvature of the geometrical wave at the ouput plane of the unstable lensguide. In other words, *in this symmetric-resonator picture the earlier apertures appear to all converge down to a single point on axis at a finite distance $L_{\infty} \equiv R_g$ behind the output plane; and this convergence point corresponds exactly to the source point of the divergent geometrical wave that characterizes the symmetric lensguide of Fig. 1(a).* In physical terms, the geometrical part of the beam in the lensguide can be described as coming from the superposition of all the most distant apertures, all of which appear to be very small and to be located to a fixed apparent distance behind the reference plane.

(Note that there are also negative-branch symmetric unstable resonator designs with overly strong converging lenses and with $M < -1$ rather than $M > +1$. In these resonators the output beam from the resonator is in fact a converging wave, and so the apparent convergence point for the distant apertures in this case will actually be in front of the reference plane.)

D. Virtual Apertures For the Canonical Unstable Lensguide

The simple symmetric model of Fig. 1(a) and Fig. 4 is particularly enlightening because it shows so clearly how the geometrical part of the output beam in an unstable resonator appears

to come from a point source at a fixed point behind the output plane. An alternative model for an unstable lensguide, used in earlier formulations of the virtual source theory, is the "canonical model" shown in Fig. 1(b). If we use this model instead, and again select a reference plane just before the output aperture, the $ABCD$ matrix going from a plane one aperture earlier to this plane is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} M & B \\ 0 & 1/M \end{bmatrix} \quad (15)$$

where B for the canonical model is related to the length L of the symmetric model by $B = L/M$. Raising this $ABCD$ matrix to the N -th power gives

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} M & B \\ 0 & 1/M \end{bmatrix}^N = \begin{bmatrix} M^N & (M^N - M^{-N})B/(M - 1/M) \\ 0 & M^{-N} \end{bmatrix} \quad (16)$$

Successive apertures seen looking back down the lensguide in this model thus appear to have *increasing* rather than decreasing effective sizes as given by

$$a_N = \frac{a}{D_N} = M^N a \quad (17)$$

and to be located at increasing distances given by

$$L_N \equiv \frac{B_N}{D_N} = \frac{M^{2N} - 1}{M - 1/M} B \quad (18)$$

This alternative mode is illustrated in Fig. 5. Note that the apparent aperture distance L_N grows considerably more rapidly with increasing N than does the aperture diameter a_N , and so the distant apertures appear to subtend a steadily decreasing solid angle. The more distant apertures in this model thus also converge to an apparent point source, located in this case at minus infinity.

The virtual source analysis can be carried out and unstable eigenmodes and eigenvalues can be calculated using either the symmetric model of Fig. 4 or the canonical model of Fig. 5, and the final results will be the same except for the presence of the geometric wavefront curvature in the output beam from the symmetric model. The symmetric model of Fig. 1(a) and Fig. 4 may, however, seem somewhat easier to visualize, and closer in spirit to the elementary geometrical picture of the unstable resonator.

III. DERIVATION OF THE POLYNOMIAL EIGENEQUATION

In this section we will use $f(x)$ to denote any one of the transverse eigenwaves characteristic of the unstable lensguide, as observed at the reference plane selected above, and $f_g(x)$ to denote the elementary spherical wave characteristic of the geometrical solution for the same resonator at the same plane. Note that there will be in general multiple lowest and higher-order examples—possibly an infinite set—of eigenwaves $f(x)$ and corresponding eigenvalues γ , but we will not add an index to distinguish different eigenwaves at this point.

Suppose then that any one particular eigenwave of the unstable system is launched into the symmetric lensguide of Figs. 1(a) and 3(a) with unit amplitude at the N -th aperture back from the output reference plane, where N is some large value, as shown in Fig. 6. Let $f^{(N)}(x)$

FIGURE 5. Apparent apertures for an unstable lensguide as seen using the canonical model.

FIGURE 6. The output wave at plane $n = 0$ consists of a geometrical wave coming from earlier $n = N$, plus edge waves from all the intervening apertures.

denote the wave pattern of this same eigenwave when it reaches the output aperture. From the physical model of Figure 6, this output wave $f^{(N)}(x)$ will consist of the geometric wave $f_g(x)$ produced by the input wave $f(x)$ coming through the distant aperture located N periods back,

plus edge waves $F_n(x)$ produced by each of the intervening N aperture edges seen looking back from the output plane. We assume therefore that we can write the output wave (for symmetric eigenmodes) in the form

$$f^{(N)}(x) = \gamma_g^N a_g f_g(x) + \sum_{n=1}^N a_n F_n(x) \quad (19)$$

The first term in this expansion is the geometric part of the solution which we assume was launched with an initial amplitude a_g (to be determined) at the aperture N periods back. This geometrical wave as it radiates out and spreads will be attenuated by the geometrical attenuation factor $\gamma_g \equiv \sqrt{1/M}$ for each period it passes through, or a total attenuation of γ_g^N by the time it reaches the output plane.

The functions $F_n(x)$ represent the edge-wave contributions at the output plane produced by a wave of unit amplitude striking the edges of each of the apertures seen looking back from the output plane, with the apertures numbered from $n = 1$ to $n = N$ looking back as in Fig. 6. A derivation of these edge-wave functions for slit and circular apertures using the Fresnel approximation is given in Appendix A. Now, since the actual wave passing down the lensguide and illuminating the successive apertures is one of the system eigenwaves, the actual amplitude striking an aperture edge n periods back will be just the output function $f^{(N)}(x)$ increased by the ratio γ^{-n} . Hence the actual amplitude a_n of each edge wave at the earlier aperture will be given by

$$a_n = \gamma^{-n} f^{(N)}(a) \quad (21)$$

and Eq. (19) becomes

$$f^{(N)}(x) = \gamma_g^{-N} a_g f_g(x) + f^{(N)}(a) \sum_{n=1}^N \gamma^{-n} F_n(x) \quad (22)$$

for the eigenfunction at the output plane.

Suppose next that we launch the same eigenwave, with the same input amplitude, except we launch it $N + 1$ periods back. By the same arguments the output wave is now given by

$$\begin{aligned} f^{(N+1)}(x) &= \gamma_g^{N+1} a_g f_g(x) + f_{N+1}(a) \sum_{n=1}^{N+1} \gamma^{-n} F_n(x) \\ &= \gamma f^{(N)}(x) = \gamma \gamma_g^N a_g f_g(x) + f^{(N)}(a) \sum_{n=1}^N \gamma^{1-n} F_n(x) \end{aligned} \quad (23)$$

Note that the second line in this equation must be true because $f^{(N+1)}(x)$ is, by definition, just the same eigenwave as $f^{(N)}(x)$ but propagated one period further along the lensguide.

By comparing the summations in the first and second lines of Eq. (23), we can see that the individual $F_n(x)$ terms in the first and second lines are automatically matched in amplitude for $n = 1$ to $n = N$, since $f^{(N+1)}(a) \equiv \gamma f^{(N)}(a)$. Equating the remaining terms on the right-hand sides of the first and third lines then gives

$$(\gamma - \gamma_g) \gamma_g^N a_g f_g(x) = \gamma^{1-N} f^{(N)}(a) F_{N+1}(x) \quad (24)$$

Both sides of this equation will have the same x dependence, as required, because the distant edge-wave term $F_{N+1}(x)$ will in fact have exactly the same form as the geometrical wave term $f_g(x)$ for a large enough value of N (i.e., for a distant enough and small enough aperture). Equation (25) thus reduces to

$$\bar{a}_g f_g(x) = \frac{(\gamma \gamma_g)^{-N}}{\gamma - \gamma_g} F_{N+1}(x) \quad (25)$$

and this expression serves to scale the amplitude a_g of the initial geometric wave launched through the distant aperture.

Putting Eq. (25) into Eq. (24) and setting $x = a$ then gives the polynomial equation

$$(\gamma - \gamma_g)\gamma^N = F_{N+1}(a) + (\gamma - \gamma_g) \sum_{n=1}^N F_n(a) \gamma_g^{N-n} \quad (26)$$

Since the F_n 's are known functions, and the eigenwave amplitude $f(a)$ has cancelled out, this becomes an $(N + 1)$ -th order polynomial equation which can be used to determine $N + 1$ eigenvalues γ for the eigenwaves in the lensguide. This is in fact the same polynomial derived by Horwitz and later authors using more mathematical and less physical arguments.

The analysis must be slightly modified for odd-symmetry modes in the strip resonator case. In this case the geometric term must be written as

$$(27)$$

and the polynomial equation becomes

$$(28)$$

This result was also pointed out by Southwell.

We inquire finally how large the order of this polynomial must be to obtain accurate results. In physical terms this is the same as asking how far back along the chain of apertures one must go until the apparent apertures become so small that their two edges effectively merge on axis, so that the edge waves from the N -th aperture back will have essentially the same form as the geometrical wave at the output plane. The phase front for the geometrical wave has the form

$$f_g(x) \propto \exp\left[-j \frac{\pi x^2}{L_\infty \lambda}\right] \quad (29)$$

while the phase fronts from the two edge waves of aperture N will have the form

$$F_N(x) \propto \exp\left[-j \frac{\pi}{L_N \lambda} (x \mp a_N)^2\right] \quad (30)$$

Suppose enough constant phase shift is added to one of these waves to make these two wavefronts coincide at one edge ($x = +a$) of the output aperture. The phase difference between the two waves at the other edge ($x = -a$) will then be given by

$$\Delta\phi \approx \frac{\pi}{L_N \lambda} a a_N \quad (31)$$

and we desire that this phase shift be a small fraction of a cycle, or $\Delta\phi \leq \epsilon 2\pi$ with $\epsilon \ll 1$. Using the formulas for a_N and L_N for large N then leads to the criterion

$$N \geq \ln \left(\frac{4\epsilon M^{1/2} F_{eq}}{M+1} \right) / \ln M \quad (32)$$

If we set $\epsilon = 1/125$ this is not significantly different from Horwitz's original criterion requiring that $N \geq \ln(250 F_{eq}) / \ln(M)$.

IV. BIORTHOGONALITY OF UNSTABLE LENS GUIDE MODES

- Typical results for eigenvalues and eigenwaves calculated using the virtual source symmetric-resonator approach—results to be supplied.

V. BIORTHOGONALITY OF UNSTABLE LENS GUIDE MODES

The mathematical theory of unstable resonators predicts that the resonator eigenmodes, call them now ϕ_n for different mode numbers n , rather than being power orthogonal to each other, should be biorthogonal to a set of transposed modes u_n which are simply the eigenmodes that would travel along the same lensguide flipped in the opposite direction. While there is little doubt as to the correctness of this prediction, the shapes of the eigenmodes in the unstable resonator case are sufficiently complex that the validity of these biorthogonality properties is not immediately apparent. Verifying the biorthogonality properties for a set of numerically calculated eigenmodes can thus serve, not so much as a test of the biorthogonality principle, as a check on the correctness of the analytical formulation and the numerical procedures.

We have therefore carried out such a check for a representative set of unstable-resonator eigenfunctions calculated using the virtual source approach in both the strip and circular mirror cases. Figure 7 shows the lowest and several higher-order even eigenfunctions calculated for a typical strip resonator, using the parameter values ...

In the symmetric case the transposed mode u_n is simply the eigenmode ϕ_n truncated by the aperture and passed through the diverging lens, so that it has a geometrical radius of curvature given by

(33)

In these calculations the transposed functions were first individually power-normalized so that $\int u_n(x) u_n^* dx = 1$ and the eigenfunctions then cross-normalized so that $\int u_n(x) \phi_n(x) dx = 1$. Table 1 then shows both cross-power and biorthogonality products of each of the eigenmodes n with other modes n .

- Results for even and odd modes??
- Self- and cross-powers??
- Variations with
- Ability to calculate field patterns at other planes using Eq. ().

VI. EXCESS NOISE FACTORS FOR UNSTABLE RESONATORS

- Text to be supplied later.

VII. CONCLUSIONS

APPENDIX A. VIRTUAL SOURCE TERMS FOR SLIT AND CIRCULAR APERTURES

- Derivation and summary of virtual source terms for circular apertures.

REFERENCES

1. A. E. Siegman, "Unstable optical resonators for laser applications," *Proc. IEEE* **53**, 277 (March 1965).
2. A. E. Siegman and R. Arrathoon, "Modes in unstable optical resonators and lens waveguides," *IEEE J. Quantum Electron.* **QE-3**, 156-163 (April 1967).
3. A. E. Siegman and H. Y. Miller, "Unstable optical resonator loss calculations using the Prony method," *Appl. Opt.* **9**, 2729-2763 (December 1970).
4. A. E. Siegman, *Lasers*, (University Science Books, 1986).
5. A. E. Siegman, "A canonical formulation for multi-element unstable resonator calculations," *IEEE J. Quantum Electron.* **QE-12**, 35 (January 1976).
6. Sanderson and Streifer, Oct. 1969, Nov. 1969.
7. D. B. Rensch, 1974.
Rogers and Erkkila, 1983.
L. A. Weinstein, *Open Resonators and Open Waveguides* (Golem Press, Boulder, Colorado, 1969).
Chen and Felsen, 1973.
Walls and McNeil, 1976
E. A. Sziklas and A. E. Siegman, "Mode calculations in unstable resonators with flowing saturable gain. I. Hermite-gaussian expansion," *Appl. Opt.* **13**, 2775-2792 (December 1974).
A. E. Siegman, *Lasers* (University Science Books, 1986).
P. Horwitz, "Asymptotic theory of unstable resonator modes," *J. Opt. Soc. Am.* **63**, 1528-1543 (December 1973).
P. Horwitz, "Modes in misaligned unstable resonators," *Appl. Opt.* **15**, 167-178 (January 1976).
G. T. Moore and R. J. McCarthy, "Theory of modes in a loaded strip confocal unstable resonator," *J. Opt. Soc. Am.* **67**, 228-241 (February 1977).
R. R. Butts and P. V. Avizonis, "Asymptotic analysis of unstable laser resonators with circular mirrors," *J. Opt. Soc. Am.* **68**, 1072-1078 (August 1978).

W. H. Southwell, "Asymptotic solution of the Huygens-Fresnel integral in circular coordinates," *Opt. Lett.* **3**, 100-102 (September 1978).

W. H. Southwell, "Virtual-source theory of unstable resonator modes," *Opt. Lett.* **6**, 487-489 (October 1981).

W. H. Southwell, "Unstable-resonator-mode derivation using virtual-source theory," *J. Opt. Soc. Am. A* **3**, 1885-1891 (November 1986).

CAPTIONS FOR FIGURES

FIGURE 1. (a) Symmetric and (b) canonical forms for a general one-aperture unstable resonator or lensguide.

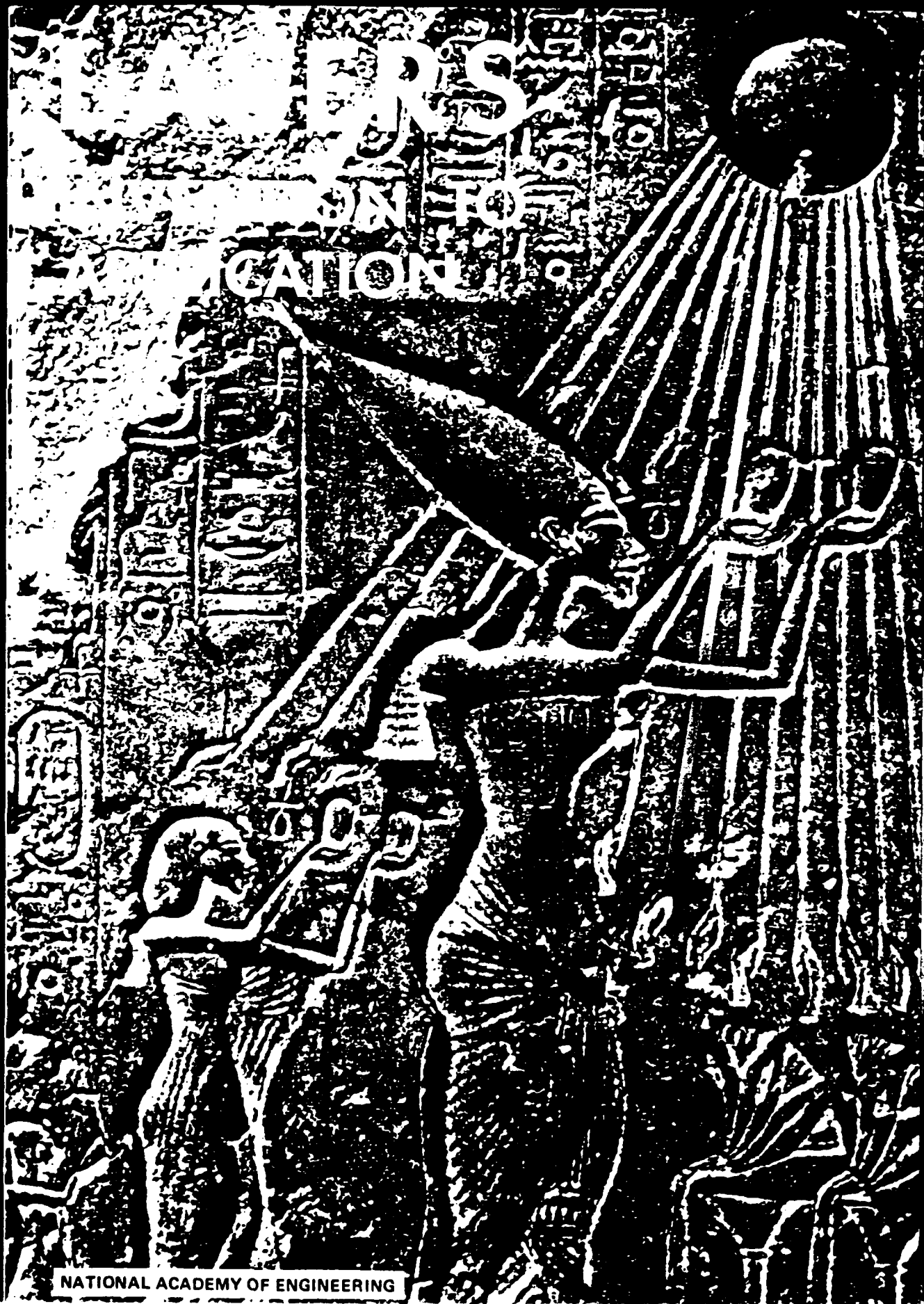
FIGURE 2. A source point x_1 as seen through an arbitrary $ABCD$ system appears to be transformed to transverse displacement x_1/D at distance B/D .

FIGURE 3. Looking backward to the N -th aperture in a symmetric unstable lensguide.

FIGURE 4. Apparent apertures seen looking backward down a symmetric unstable lensguide.

FIGURE 5. Apparent apertures for an unstable lensguide as seen using the canonical model.

FIGURE 6. The output wave at plane $n = 0$ consists of a geometrical wave coming from earlier $n = N$, plus edge waves from all the intervening apertures.



MEMBERS CONTRIBUTION TO COMMUNICATION

NATIONAL ACADEMY OF ENGINEERING

The Laser: Still Young at 25?

Anthony E. Siegman

The first laser device was operated just over 25 years ago. In the subsequent two-and-a-half decades, laser devices of many diverse types have produced an enviable record of accomplishments in fundamental science, applied technology, medicine, and even home entertainment. Although not yet of the economic importance of conventional electronics, the laser industry is significant and growing.

This paper will provide a brief review of the history of the laser; describe some of the characteristics and performance capabilities of different types of laser devices; give a short introduction to the breadth and diversity of laser applications; and finally, summarize some of the exciting current accomplishments and possible future advances in laser technology.

LOOKING BACK 25 YEARS

To begin with the very early history or, perhaps, the mythology of the laser, the following passage from H. G. Wells's famous 1896 novel *The War of the Worlds*, about an invasion of Earth by Martians, gives a reasonably accurate description of how one might make and then use a laser.

In some way they [the Martians] are able to generate an intense heat in a chamber of practically absolute nonconductivity. . . . This intense heat they project in a parallel beam against any object they choose, by means of a polished parabolic mirror of unknown composition. . . .

The second part of this quotation then says:

However it is done, it is certain that a beam of heat is the essence of the matter. What is combustible flashes into flame at its touch, lead runs like water, it softens iron, cracks and melts glass, and when it falls upon water, that explodes into steam. . . .

To anyone who has seen the materials processing effects produced by even a medium-power laser beam, Wells's description, written at the turn of the century, will seem a remarkably accurate account of the effects of the beam from a modern carbon dioxide (CO_2) laser of perhaps a few kilowatts of power output. It should also come as no surprise that we now know, from our own space probes, that the natural atmosphere of the planet Mars consists primarily of carbon dioxide, and that, in fact, natural laser action, pumped by sunlight, occurs in the Martian atmosphere.

We of course do not really believe that Martian invaders brought CO_2 lasers with them to Earth a century ago, but the basic principles of stimulated emission from atoms or molecules were first recognized by Einstein, Ladenburg, and others in the 1930s. The first man-made device to use these principles came only in 1955, when Charles H. Townes of Columbia University operated the first ammonia beam *maser*—the acronym coined by Townes to stand for *microwave amplification by stimulated emission of radiation*. This was closely followed by similar developments by N. G. Basov and A. M. Prokhorov in the Soviet Union. Townes's first maser was, of course, not an optical device, but a weak microwave oscillator at 24 GHz.

In the years that followed, many researchers gave much thought to the possibility of optical masers, or lasers. This process of development was much assisted by the concept of the continuously pumped three-level microwave maser developed by N. Bloembergen at Harvard University in 1956. In 1958 Townes and Arthur L. Schawlow of Bell Laboratories published a paper and patent application that gave a theoretical recipe for laser action. The first successful optical-frequency laser device, the pulsed ruby laser, was actually developed in 1960 by an industrial researcher, Theodore H. Maiman, in the Hughes Research Laboratories in Malibu, California.

Maiman produced this first laser action by placing silver mirrors directly on the end of a synthetic crystal of ruby, and then pumping or exciting this crystal with an intense flash of light from a standard photographic flash lamp. Maiman's pioneering advance was rapidly followed by the development of a number of other laser devices by the IBM and Bell Telephone Laboratories; in particular, the first continuously running, electrically pumped gas lasers were developed in the same year at



FIGURE 1 Arthur L. Schawlow demonstrating that a flash of red light from a small ruby laser breaks the dark-colored inner balloon without damaging the transparent outer balloon. This procedure exactly mimics the way laser light can be used to repair a detached retina inside the eye-ball, make a weld inside a closed vacuum chamber, or trigger a chemical reaction inside a closed chemical cell. Photograph by Frans P. Alkemade.

Bell Labs. The laser field has been characterized ever since by the continual emergence of new and ever-surprising laser systems, a process that is still going on today.

Schawlow, one of the most distinguished scientists in the laser field, was a corecipient of the 1981 Nobel Prize for his work on laser spectroscopy. He was also the pioneer of the first edible laser: a glass cell full of unflavored gelatine, which he first operated as a laser and then ate. Figure 1 shows another of Schawlow's distinctive demonstrations.

HOW LASERS WORK

The construction of a laser begins with a collection of atoms. These atoms or molecules can be gaseous, liquid, or solid in form, but they are characterized, as are all atoms, by a set of discrete and distinctive quantum energy levels.

Next, some form of pumping process must be applied to these atoms. This pumping process can be accomplished in a great



many ways. In all cases, however, its essential function is to excite or lift some of the laser atoms out of their lowest quantum energy level and into upper energy levels (see Figure 2).

If atoms can be excited into upper energy levels and, more importantly, a condition of population inversion can be achieved, in which more atoms are excited into some upper atomic level than into some lower atomic level, then laser action can occur. If a beam of light tuned to the transition frequency between those two levels is sent through the collection of atoms, that light beam will be amplified through the process of stimulated emission. Stimulated emission means simply that the electromagnetic fields in the light beam cause the atoms to drop down from the more heavily populated upper level into the less heavily populated lower level, giving up their energy to the light beam in the process, in phase (i.e., coherent) with the exciting field.

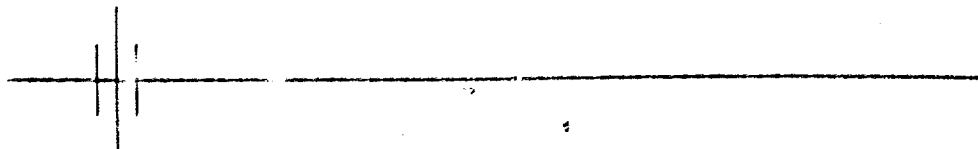
Next, a carefully aligned laser mirror is added at each end of the collection of atoms to form an optical resonator, in which the lightwave can bounce back and forth many times. If the round-trip gain in this resonator exceeds the round-trip losses due to absorption and finite mirror reflectivity, then the light signal in this laser cavity can build up to a coherent optical oscillation, exactly the same as in an audio frequency oscillator or a radio frequency transmitter.

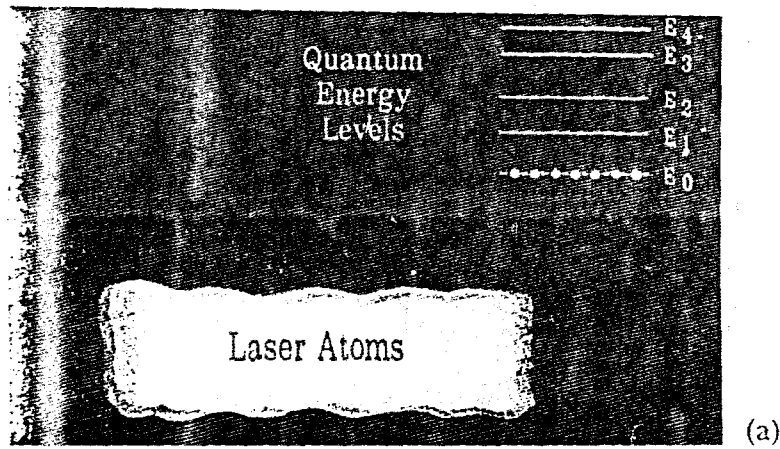
The optical signal in this laser cavity can be highly monochromatic, or of a single frequency, or temporally coherent because it is a true coherent oscillator. It can also be highly directional (collimated) or spatially coherent because of the directional control produced by the two mirrors.

THE PRESENT STATUS OF LASERS

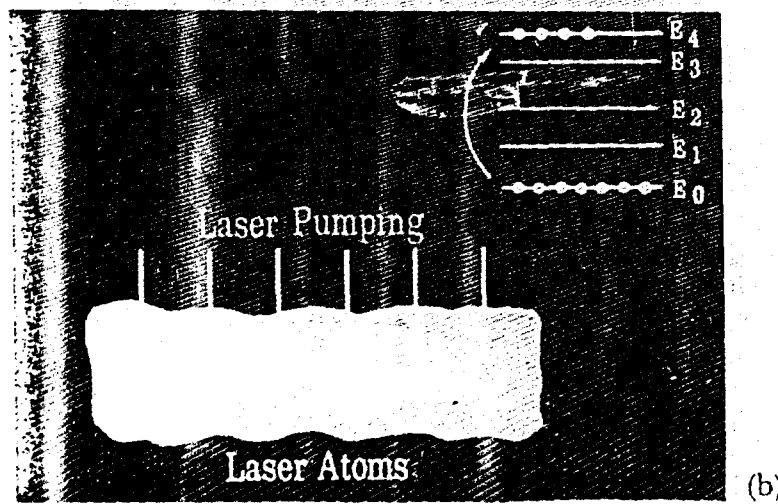
An enormous number of widely different types of laser devices have now been discovered (Table 1). They cover the wavelength range from shorter than $1,000 \text{ \AA}$ in the vacuum ultraviolet to longer than $800 \text{ }\mu\text{m}$ in the far infrared or millimeter wavelength range. The list of laser materials covers essentially every form of matter, from simple gases and solids to liquids, plastics, flames, jet engine exhausts, and interplanetary space. Indeed, Schawlow's law, yet to be disproved experimentally, says that anything will lase (i.e., generate a beam of laser light) if it is hit hard enough. Of course, if something does not lase, then it was not hit hard enough.

The list of atoms in which laser action has been obtained

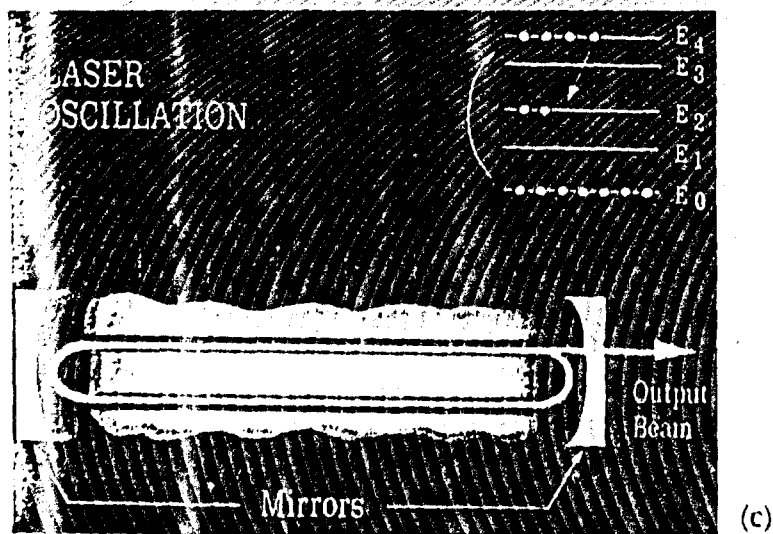




(a)



(b)



(c)

FIGURE 2 (a) A collection of laser atoms and their quantum energy levels. (b) The laser pumping process. (c) Stimulated emission and laser oscillation.

TABLE 1 Present Status of Lasers

Laser Characteristic	Present Status
Wavelength range	From longer than 800 μm to shorter than 150 \AA
Laser materials	Gases: atoms, ions, molecules, excimers Liquids: organic dyes, H_2O solutions, Scotch whisky Solids: crystals, glasses, plastics, semiconductors, plasmas, flames, jet engine exhausts, interstellar space, planetary atmospheres
Laser atoms	More than 100 individual atoms and ions, innumerable molecular species
Laser transitions	More than 10^6 discrete laser lines
Peak powers	Greater than 10^{13} W
Average powers	Greater than 1 MW
Frequency stability	Few parts in 10^{14}
Pulsewidths	Less than 8 fs (8×10^{-15} s)
Tuning ranges	Greater than 200 \AA (about 24,000 GHz)

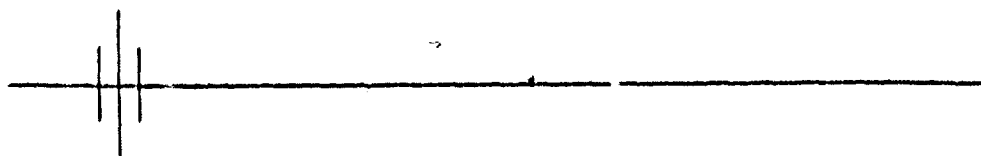
covers essentially the entire periodic table, in both atomic and ionized forms, in addition to a virtually unlimited list of molecular species. The number of individual laser transitions is practically uncountable—there are, for example, more than 200 individually identifiable laser transitions between different quantum energy level pairs in the neutral neon atom alone. Therefore, the number of possible laser transitions is almost surely in the millions.

AN ELECTRONIC OLYMPIC GAMES?

One way of dramatizing the extraordinary capabilities of lasers might be through a sort of "Electronic Olympic Games," a set of competitive events to see which electronic devices—transistors, integrated circuits, vacuum tubes, or lasers—set the current performance records in generating the highest powers, the shortest pulses, the greatest frequency stability, the lowest noise figure, and other limits. How many gold medals might the laser, in particular, win in such a competition?

FREQUENCY RANGE, TUNING RANGE, AND BANDWIDTH

Laser devices of many different kinds will clearly win all the available gold, silver, and bronze medals for frequency range, tuning range, and instantaneous bandwidth. The frequency or wavelength range over which different kinds of lasers and



masers can operate extends from the subaudio to the x-ray regions.

The fractional tuning range of most individual lasers is relatively small, limited by the atomic linewidth of the laser transition used. In absolute terms, however, the bandwidths or tuning ranges of many lasers still extend over tens to hundreds of gigahertz. So-called organic dye lasers, along with semiconductor lasers, can, in fact, be continuously tuned over linewidths of hundreds of angstroms; commercially available dye lasers with multiple dyes can be tuned continuously over essentially the entire visible and near-infrared spectral range. The applications of such lasers in chemistry and chemical diagnostics can be easily imagined.

Without going into further detail here, it should also be noted that an ordinary dye laser with a 200-Å spectral width has a frequency bandwidth sufficient to transmit the equivalent of one simultaneous telephone channel for every person on earth.

PEAK POWER

For setting peak power output records, laser devices also stand absolutely supreme, in large part because of their ability to generate very short pulses. Indeed, rather modest mode-locked lasers of tabletop size can easily produce optical pulses with instantaneous peak optical powers in excess of 10^{13} W, or several times the total installed electrical generating capacity of the United States—though these pulses last for only a few trillionths of a second.

These same laser pulses can then be focused into spots only a few optical wavelengths in diameter to produce peak power intensities of billions of watts per square centimeter—sufficient to tear atoms apart, break molecular bonds, produce intense nonlinear optical effects, and melt and vaporize any material. Even a small pulsed solid-state laser, for example, can readily drill or cut through steel, ceramic, diamond, or any other material.

CONTINUOUS POWER OUTPUT

The continuous or average powers available from certain laser devices are also impressive, although it is uncertain whether they exceed those available from all other high-power electronic devices, including klystrons, high-power triodes, and even motor generator sets. Of particular interest here, however, is the enormous diversity of pumping or excitation methods that can

be used for producing laser action (Table 2). These include not merely electrical discharges of all kinds and optical pumping methods using almost any conceivable light source but also laser action in flames, plasmas, chemical reactions, focused sunlight, nuclear reactions, and nuclear explosions. Especially striking is the fact that several types of natural laser and maser action also occur (without mirrors) both in interstellar space and in planetary and solar atmospheres.

Particularly impressive in the context of high-power lasers are the chemical and gas-dynamic laser systems, which can convert the energy of a chemical reaction, or pure heat energy in gases, directly into coherent laser radiation with extremely high energy output. Figure 3 shows, for example, a large gas-dynamic laser built in the early 1970s. This laser burned a chemical fuel (cyanogen) in the lower chamber, and then sent the hot gases upward through supersonic expansion nozzles into the laser region to produce some hundreds of kilowatts or more of laser power. The water-cooled mirrors and mirror mounts are in the boxes at the end, and the hot gases are exhausted through the deflectors at the top. This is one of the few lasers that must be bolted down because it has thrust.

The rule of thumb for such chemical lasers is that the combustion of 1 kg of fuel can typically produce a sufficient number of excited molecules to provide several hundred kilojoules of coherent (though multiwavelength) laser output energy. A fuel supply of a few kilograms per second suffices, therefore, to power a 1-MW laser oscillator. Israeli scientists have

TABLE 2 Laser Pumping Methods

Pumping Method	Laser Action
Optical pumping	Laser materials pumped by flash lamps, arc lamps, tungsten lamps, exploding wires, light-emitting diodes, flames, focused sunlight, other lasers
Gas discharges	Direct electron and collisional excitation in glow discharges, arc discharges, hollow cathode discharges
Chemical reactions	Laser action following chemical mixing, flash photolysis, flame photolysis
Direct electrical	Direct electrical excitation in semiconductor injection lasers, electron beam-pumped solids and gases
Gas-dynamic lasers	Laser action derived from hot gases, supersonic expansions, shock fronts
Plasmas	Laser action in plasma pinches, laser-induced plasmas
Nuclear reactions	Fission fragment pumping of gas lasers
Nuclear explosions	Atomic bomb-pumped x-ray lasers
Natural lasers	Sunlight, interstellar radiation, particle beams

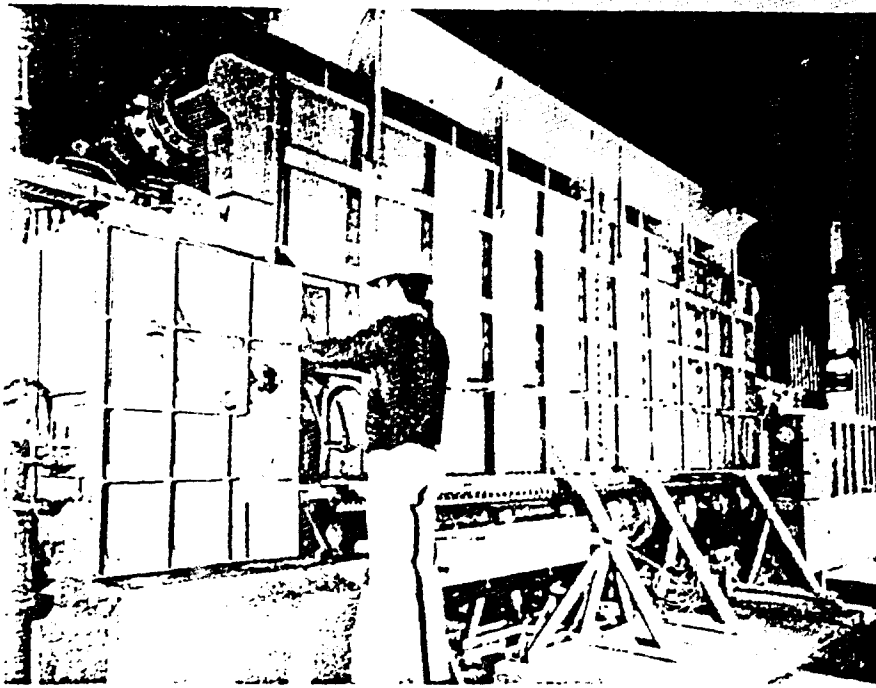


FIGURE 3 A large high-power gas-dynamic laser.

even developed a gasoline-fueled gas-dynamic laser, with the initial fuel mixture being ignited by an automobile spark plug.

EFFICIENCY

To be conservative, the laser should probably receive only a silver medal in the category of average power. Many commonly used lasers are also, unfortunately, far less efficient in the use of electrical input energy than would be desirable. The common small helium-neon laser has a typical operating efficiency of only a small fraction of a percent, although there are also useful types of gas and semiconductor lasers that have efficiencies of 60–70 percent from direct electrical input. In this category the laser would receive a bronze medal.

PULSEWIDTH

In the competition for generating the shortest possible pulses, however, the laser is second to none. The units for expressing the duration of a pulse scale downward in jumps of 1,000 from seconds to milliseconds, microseconds, nanoseconds, picoseconds—with 1 ps already shorter than any form of electronics can go—and finally down to femtoseconds, or units of 10^{-15} s.

One of the most astonishing recent accomplishments in laser technology has been the generation of fully coherent optical pulses as short as 8–10 fs. These have provided truly extraordinary new scientific capabilities for exciting, probing, and measuring internal processes in atoms and molecules, chemical reactions, biological processes, and solid-state physics—far beyond the time resolution that can be achieved electronically, now or in the near future.

FREQUENCY STABILITY AND SPECTRAL PURITY

Lasers may also be judged with respect to absolute frequency stability and spectral purity. For many decades the international standard for measurements of length has been not the historic meter bar but a visible wavelength derived from an incoherent, or nonlaser, gas-discharge light source. The international standard of time has been a microwave atomic clock, which is not quite but is almost a maser.

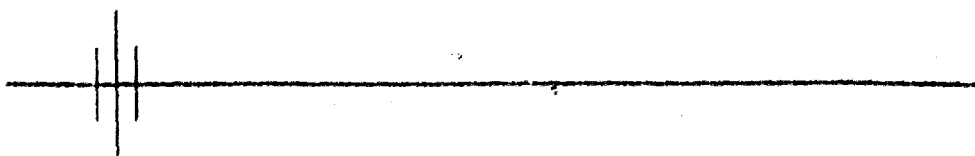
Certain laser oscillators can, however, have a spectral purity and absolute frequency stability so extraordinary that just in the past few years the international standards for both frequency and time have been, by definition, unified into a single laser device. That is, the new standards for both distance and time are now provided by a certain ultrastable laser oscillator transition located in the middle infrared region of the spectrum.

It is more than a little disconcerting to realize what this means: The goal of Albert A. Michelson and so many other distinguished physicists in the past—to make ever more precise measurements of the velocity of light—will no longer even be meaningful. Now that the basic standards of time and length are one and the same laser transition, the velocity of light, c , is reduced to a mere defined quantity: the numerical relationship between one wavelength and one period of this laser frequency. One can never again measure c .

ANTENNA BEAMWIDTH

The antenna properties or the antenna beamwidth of a laser beam are also important. The beam traveling back and forth inside a laser cavity is extraordinarily parallel or well collimated, and the beam outside the laser retains these extremely directional properties, spreading only slightly as it propagates (see Plate 1).

Such a laser beam thus provides a kind of “weightless string” that neither sags nor blows in the wind and so is extremely useful



as an alignment tool for building construction, tunneling, pipe laying, and many other civil engineering works. This is, in fact, one of the simplest but most significant commercial applications for the laser, and the fortunes of some laser companies in the past have risen and fallen with the construction market.

In technical terms, the beam angle in radians for a collimated electromagnetic wave coming from an antenna is more or less equal to the number of wavelengths across the antenna aperture. For a visible laser beam and a diffraction-limited, 10-cm-diameter aperture—which is not at all difficult to achieve in practice—this means a beamwidth of 10 microradians; this means, in turn, that a laser easily has the ability to illuminate a spot not much more than a mile wide on the face of the moon. A microwave antenna with the same beamwidth would have to be several kilometers in diameter.

At present, laser antennas up to a meter in diameter are commonplace, for example, in satellite and lunar ranging systems. Given the narrow beams and high peak powers of lasers, laser radar echoes are routinely obtained from optical reflectors located on the moon, with accumulated range accuracies of a few centimeters in the distance to the moon. By making such laser-ranging measurements to a cooperative orbiting satellite simultaneously from multiple stations, it is possible first to determine the satellite orbit and its perturbations with great accuracy and then to determine the relative positions of the laser stations on earth, and thus to map the earth with comparable accuracy even across seas and oceans.

COMPUTING CAPABILITY

Laser light certainly will be—already is—of enormous importance in fiber-optic communications generally, including fiber-optic computer networks and communications between and within computers. Less well established, however, are the pure computing capabilities of lasers—the so-called photonic logic possibilities that may come in the future.

For sheer power in general computing, all the gold medals will probably continue to go to the silicon chip and its derivatives. The silver medals may well go to gallium arsenide or other new forms of ultrafast electronics rather than to any kind of "photonic computers."

Lasers can offer unique capabilities in certain specialized techniques that are more or less computational in nature, such as holography and certain kinds of image processing. However, the laser would receive at most a bronze medal in this area.

Out of a dozen or so major Olympic events, therefore, laser devices of various kinds will win at least seven or eight gold medals, half a dozen silver, and many bronze—a record that no other class of electronic devices can approach.

LASER APPLICATIONS

To date, the applications of the laser in all fields of scientific measurement have been diverse, extraordinary, and unique. Measurements have been and are being made that simply could not be made in any other fashion. Chemistry, biology, and physics laboratories use many types of lasers to probe, measure, and modify the fundamental properties of matter. Mechanical and aeronautical engineering laboratories are equally well equipped with lasers. For example, laser beams are projected into huge wind tunnels to measure local flow velocity and turbulence.

Lasers are extraordinary tools for identifying materials as well. A medical researcher, by focusing a weak laser beam on an experimental object, can painlessly vaporize a minuscule sample of animal or human tissue for spectroscopic diagnosis of atomic composition, trace elements, and other components. Similarly, an archaeologist can vaporize a tiny sample of a suspect artifact to see if its chemical makeup agrees with its alleged origin.

Outside of pure science and engineering research and development, however, lasers have also been applied in equally diverse and often unanticipated ways in many fields of commerce, manufacturing, industry, and medicine.

Lasers are used in bar code scanners for retail stores, inventory control in warehouses, and supermarket checkout stands. Lasers are found in nearly every elementary surveying instrument these days—for example, lasers mounted on transits are used to obtain contours for leveling of rice paddies and to control dredging barges and automated bulldozers. Laser ranging instruments are used for highway surveying and housing subdivision construction.

In industry, laser beams offer both measuring and manufacturing tools that are flexible and versatile; can be precisely controlled; are well adapted to robotics and numerical control; can be adapted to almost any material or environment; are clean, reliable, and economical to run; and never grow dull.

The simplest industrial applications of high-power lasers come, of course, in the straightforward cutting of materials. Examples include the cutting of armor plate; the cutting of cast iron, without heating or annealing the surrounding material; the cutting of complex patterns in plywood, glass, plastic, or

cardboard; and the cutting of cloth for clothing, under computer control, with self-sealing of the fabric edges and minimal waste of material.

But beyond this, there is laser heat treating; laser surface hardening; laser scribing, annealing, deburring, soldering, and resistor trimming; the cutting and repairing of integrated circuits; and the drilling of precise holes in turbine blades (Plate 2), as well as in plastic irrigation pipes and rubber baby bottle nipples.

The laser has also already truly revolutionized the document-scanning, typesetting, newspaper platemaking, and printing industries, at both the high and low ends of the scale. Today, one can print either a newspaper or an interoffice memo with all the information, including the typefaces, stored in a computer memory and printed out by a computer-controlled scanning laser beam. Similar concepts can also be used for the fast and easy marking and labeling of complex mechanical parts made from almost any material. Lasers are now appearing even in consumer products, such as laser video recorders and audio compact disc players.

All of these commercial, industrial, and home applications of laser devices, however diverse and ingenious, thus far have been only preliminary. Indeed, the worldwide sales of laser devices in 1985 totaled only about \$400 million—\$500 million—or about a quarter of the sales of small computers in the same period by Apple Computers alone.

The worldwide sales of laser systems for all purposes—scientific, engineering, industrial, medical, and military, including ancillary equipment, controls, and materials handling—is believed to have totaled about \$4 billion—\$5 billion in 1984. Both of these sales figures have been growing annually by 30 percent or more in recent years, although with much higher growth rates in some areas.

In the opinion of most observers, the real boom in the use of lasers in industrial and manufacturing processes is just beginning. For example, large machine tool manufacturers and small laser companies are only now beginning to merge on a wide scale.

Beyond industrial applications, one of the fastest growing areas of laser application is in medicine. Laser surgery is now being used not only to remove skin tumors and conduct other external surgery and to treat many eye diseases but also in ear and throat surgery and gynecology. In addition, using fiber-optic delivery systems, laser surgery is done even inside the intestinal tract and blood vessels.

A form of cancer therapy with lasers is also the subject of much investigation and hope in the medical field. In this



therapy, laser light of the proper wavelength activates a photosensitive chemical within malignant human tissue, releasing singlet oxygen that destroys the cancer cells.

There are, of course, also many military applications of the laser. Some of the most successful include laser-guided bombs; laser communication links, and laser range finders and aiming devices for guns.

NEW DEVELOPMENTS

Lasers of all types have proved extraordinarily useful devices for bettering the human condition in nearly every area of life, and they will become even more so in future years. But beyond these practical applications, fundamental new basic research advances in the laser field are, even after 25 years, still emerging nearly as rapidly as in the laser's early years.

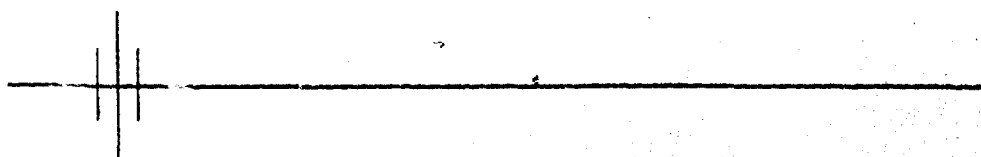
FREE ELECTRON LASERS

The last few years have seen, for example, the emergence of the so-called free electron laser, in which coherent oscillation at visible or infrared wavelengths is generated by passing the beam from a high-quality electron accelerator through a suitable wave-propagating structure. This device is not really a laser at all, but it is nonetheless a revolutionary development in coherent optical sources whose capabilities are still only in the infant stage.

These devices can provide marvelously tunable, efficient, high-power sources in the far infrared region, from 50 μm out to a few millimeters in wavelength, where laser sources are still somewhat limited. Scientific applications of such sources are manifold; unforeseen industrial applications are equally certain to emerge. The free electron laser may, with further development, also provide a similarly useful source in the visible and ultraviolet regions of the spectrum.

FEMTOSECOND OPTICAL PULSES, BISTABILITY, CHAOS, AND SOLITONS

The incredible advances made in femtosecond laser pulses during the past few years represent one example of progress in the field. Other fundamental developments have also come within the past few years in the basic understanding of new concepts of optical bistability, instabilities, and chaotic optical



behavior in lasers. Unexpected developments also occurred in optical solitons and soliton lasers that emerge when such pulses propagate through optical fibers.

TRAPPING AND COOLING OF SINGLE ATOMS AND IONS

Laser researchers are only now succeeding in trapping individual atoms and ions, or small clouds of atoms, and then cooling them to temperatures in the millikelvin range. Once trapped in this fashion, these atoms can be examined and interrogated with laser beams in ways never before possible. This will extend the ultimate precision of physical measurements and laser standards far beyond the few parts in 10^{10} that is now achievable to ultimate accuracies of a few parts in 10^{13} or 10^{14} or even better.

MULTIPLE QUANTUM WELL STRUCTURES

Another fundamental development of the past few years, important for both electronics and lasers, has been the so-called multiple quantum well structures, or artificial layered materials. It has now become possible, using several different techniques such as molecular beam epitaxy or metal-organic chemical vapor deposition, to prepare layered synthetic materials by depositing under precise control a discrete number of atomic layers, first of one material—for example, gallium arsenide—and then another—such as aluminum arsenide—in an alternating sequence.

The result is an essentially perfect artificial crystal with an adjustable period or layer thickness in the range of a few hundred angstroms. Because the properties of the electrons in these multiple quantum well structures can differ greatly from ordinary materials, the electronic and optical properties of these materials offer remarkable new capabilities, including much faster forms of conventional electronic devices.

In optics, the result has already been much more efficient and shorter wavelength diode lasers for use in fiber-optic communications or audio compact disc players, as well as improved photodetectors, light modulators, and other electro-optic devices. By using these diodes as pumps for other laser materials, ultraminiaturized lasers of many different types can be produced.

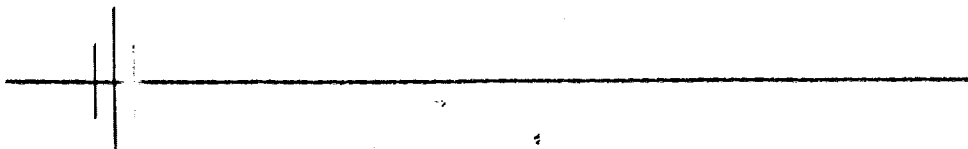
X-RAY LASERS

Another topic of continuing interest is the extension of laser techniques to the x-ray region. This will always be a difficult task

for several reasons, one of which is the difficulty of providing mirrors at these wavelengths. Another basic barrier derives from the fundamental equations of laser theory that say that obtaining laser action becomes more difficult at somewhere between the third and the fifth power of the inverse laser wavelength. Lasers in the x-ray region, if they ever become common, are thus unlikely to be similar to lasers in the optical region.

Nonetheless, Lawrence Livermore Laboratories has recently announced the observation of stimulated emission and laser amplification in the far-ultraviolet or soft x-ray region, at wavelengths of 150–200 Å, in a target plasma pumped by a high-power laser beam. The same lab has also produced a true, if short-lived, one-shot x-ray laser in the few-angstrom region by pumping a suitable laser material directly with a small nuclear explosion.

Given all these recent advances, therefore, it seems clear that the laser field truly is still young, vigorous, and exciting, even as it passes the mature age of 25.



Two-dimensional calculations using one-dimensional arrays, or "Life on the Skew"

Anthony E. Siegman

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305

(Received 13 January 1988; accepted 12 July 1988)

Two-dimensional problems with periodic boundary conditions can be reprogrammed as one-dimensional arrays with no special handling of array indices at boundary edges or corners being required. The result is physically equivalent to a kind of vertically skewed periodic boundary condition.

Two-dimensional problems with boundary conditions going to zero at infinity—for example, electrostatic field problems or two-dimensional quantum wavefunction problems—are often analyzed using periodic boundary conditions instead. That is, instead of taking the boundary conditions to be at infinity, one assumes that the problem is confined within a large rectangular box that is periodically replicated out to infinity in both transverse directions, with the primary region taken large enough so that the replicated elements in adjoining regions have little physical effect within the primary region of interest. If numerical calculations for such a problem are set up using a two-dimensional rectangular grid within the primary region, the situation then appears as in Fig. 1. The function values at correspondingly numbered points in each region are assumed to be identical, so that the physical solution also iterates periodically in both directions.

In carrying out such calculations, the dynamical or iterative behavior of the function value in a given cell is usually determined by the function values in the surrounding nearest-neighbor cells. Suppose, for example, that one stores the function values in an array $F(I, J)$ with indices $I = 1$ to M and $J = 1$ to N corresponding to the horizontal and vertical coordinates in Fig. 1. The iterative behavior for each element $F(I, J)$ will then be calculated using the values of the surrounding array elements with indices $I' = I - 1, I$, and $I + 1$, and $J' = J - 1, J$, and $J + 1$. The necessary calculations can be carried out for every element $F(I, J)$ using two nested DO or FOR loops over $I = 1$ to M and $J = 1$ to N , except for those cells or array elements located in the outside rows and columns of the rectangular region, where the special nature of the periodic boundary conditions must be taken into account.

In Fig. 1, for example, the nearest-neighbor cells immediately to the right of element $F(4, 6)$ are not cells $(3, 7)$, $(4, 7)$, and $(5, 7)$, but rather cells $(3, 1)$, $(4, 1)$, and $(5, 1)$. In general, with periodic boundary conditions the next column to the right of column J for cells in the rightmost column of the array is not $J' = J + 1$ but rather $J' = J + 1 - N$, and similar but different conditions apply for cells in the leftmost column, the top and bottom rows, and the four corners. To handle these special cases one must either test all of the I and J indices throughout the entire calculation and apply special formulas for edge and corner cells or one must evaluate all the indices using a MOD function or one must have the nested DO loops run

from $I = 2$ to $M - 1$ and from $J = 2$ to $N - 1$, and write two additional loops and four single blocks of special code to handle the outside columns, the top and bottom rows, and the four corners.

The purpose of this note is to point out a simple but general way to program these sorts of two-dimensional calculations using only a *single-index array*, with no boundary condition tests at all required during the calculation. To visualize this method, consider an alternative one-dimensional indexing scheme for the same region as shown in Fig. 2. If the cells are indexed as shown, the nearest neighbors for any array element $F(K)$ for $1 \leq K \leq M \times N$ will be the elements $K' = K \pm 1$, $K \pm N$, $K \pm 1 + N$, and $K \pm 1 - N$. This will remain true even along the outside columns of the array if the periodically replicated regions on the right and left sides are viewed as being shifted or skewed upward or downward by one cell width per replication, as shown in Fig. 2. In addition, no special handling of the top and bottom rows will be required if the basic

	8,6	8,1	8,2	8,3	8,4	8,5	8,6	8,1
1,6	1,1	1,2	1,3	1,4	1,5	1,6	1,1	
2,6	2,1	•	•	•	•	2,6	2,1	
3,6	3,1	•	•	•	•	3,6	3,1	
4,6	4,1	•	•	•	•	4,6	4,1	
5,6	5,1	•	•	•	•	5,6	5,1	
6,6	6,1	•	•	•	•	6,6	6,1	
7,6	7,1	•	•	•	•	7,6	7,1	
8,6	8,1	8,2	8,3	8,4	8,5	8,6	8,1	
1,6	1,1	1,2	1,3	1,4	1,5	1,6	1,1	

FIG. 1. A two-dimensional physical problem encoded onto an 8×6 rectangular grid with periodic boundary conditions. The eight nearest neighbors for a cell located on the right-hand boundary are enclosed within the dashed line.

42	43	44	45	46	47	48	1
48	1	2	3	4	5	6	7
6	7	8	9	10	11	12	13
12	13	.	.	.	17	18	19
18	19	.	.	.	23	24	25
24	25	.	.	.	29	30	31
30	31	36	37
36	37	42	43
42	43	44	45	46	47	48	1
48	1	2	3	4	5	6	7

FIG. 2. The same physical problem reorganized using a singly indexed 48-element array with skewed periodic boundary conditions.

array $F(K)$ for $K = 1$ to $M \times N$ is extended with $N + 1$ "folded over" elements extending from $K = 0$ down to $K = -N$ on the lower end and from $K = N \times M + 1$ up to $K = N \times M + N + 1$ on the upper end. (Variations on this array indexing for languages that do not allow negative index values can be trivially obtained.)

Suppose, for example, that the procedure used to update the value of any cell $F(K)$ in terms of its eight nearest neighbors in one iteration is denoted by $G[F, K, M, N]$. One complete updating of the entire array $F(k)$ can then be accomplished, using a temporary array $FT(K)$, by the coding

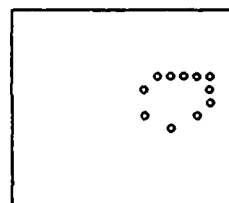
```

FOR K = 1 TO M×N
  FT(K) = G[F,K,M,N]
NEXT K
FOR K = 1 TO M×N
  F(K) = FT(K)
NEXT K
FOR K = 1 TO N + 1
  F(M×N + K) = F(K)
  F(1 - K) = F(M×N + 1 - K)
NEXT K

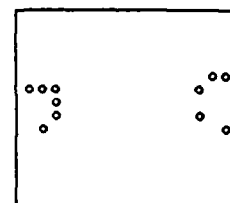
```

The third loop copies the updated values inside the primary region into the extensions at each end of the array, for use in calculating the next iteration. No testing of the index K for boundary conditions is required at any point in the calculations.

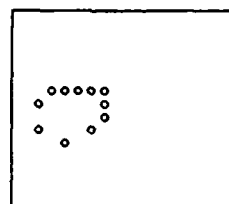
The technique outlined in this note first occurred to the author while coding the Game of Life for a microcom-



Generation 80



Generation 88



Generation 96



Generation 332

FIG. 3. The Game of Life programmed using skewed periodic boundary conditions. A "glider" leaving the grid on the right-hand side reenters the grid on the left-hand side displaced downward by one row of cells.

puter. If the Game of Life is coded in this fashion, to cite a trivial example, the result is a kind of "Life on the Skew" in which a horizontally traveling glider that flies out of the rectangular region on the right-hand side will reenter on the left-hand side vertically displaced downward by one grid unit on each transit, as shown in Fig. 3. This continues even after it reaches the bottom of the window, at which time points in the glider falling below the bottom of the window begin appearing in the top lines of the window, until the entire glider is back inside the window at the top.

The underlying concept described here is surely not new, though we have not succeeded in tracking down references to it. A few numerical analysis authorities consulted in person seem to be aware of the approach while others do not. Though the Game of Life example is quite trivial, the same approach would seem to be very useful for programming more serious physical problems. Periodic boundary conditions are often employed in numerical calculations, both to simplify the coding, and also to ensure exact conservation of energy, quantum probability, or some other physical quantity. So long as the physical size of the primary region and the number of cells are both large enough, the small "skewing" of the periodic boundary conditions that results from the single-array indexing should have negligible effect on the accuracy of the solution, with the benefits that numerous tests are eliminated and the single-array coding can be significantly simpler, especially in assembly language.

ULTRAFAST NONLINEAR OPTICAL MEASUREMENTS USING THE
TUNABLE-LASER-INDUCED GRATING TECHNIQUE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Charles Earl Barker
November 1988

ABSTRACT

The Tunable-Laser-Induced Grating technique is a nonlinear-optical frequency-domain method for the measurement of atomic and molecular processes which occur on picosecond and femtosecond timescales. This dissertation presents a continuation of the efforts on this technique begun by Trebino and Siegman.

A four-wave mixing calculation of the frequency response of the induced grating diffraction efficiency is presented. The resulting mathematical expression describes the general features of Tunable-Laser-Induced Grating spectra and provides a framework for discussing effects observed in transparent media such as the optical Kerr effect in carbon disulfide.

Experimental results on several systems of interest are also reported. In particular, we find that the use of parallel polarized excitation beams creates intensity gratings which drive long-lived scalar responses of the sample material. Examples of intensity gratings include thermal gratings in saturable absorber dyes and electrostrictively-driven acoustic waves in transparent liquids, both of which obscure the measurement of ultrafast processes taking place in these systems. By using orthogonally polarized excitation beams, we have eliminated these intensity-grating effects and have measured ultrafast responses of both the saturable absorber dye malachite green and the optical Kerr liquid carbon disulfide. High temporal resolution measurements of the optical Kerr response of carbon disulfide in particular indicate clearly the ultrafast inertial character of the nuclear component of carbon disulfide's optical Kerr response.

Finally, the Tunable-Laser-Induced Grating method measures only the square magnitude of the sample's response in the frequency domain, and the resulting loss of phase information can lead to ambiguities in determining the relative strength factors in multi-component responses. We show, both theoretically and experimentally, how the controlled introduction of a complex-valued coherent background can eliminate this so-called "rise-fall" ambiguity.